IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Crafton, Corey M. et al.	:	
		:	Examiner: Kaushal, Sumesh
		:	
Serial No.:	09/987,763	:	Art Unit: 1633
		:	
Filing Date:	November 15, 2001	:	
		:	

Entitled: NUCLEOTIDE SEQUENCES FOR TRANSCRIPTIONAL REGULATION OF CORYNEBACTERIUM GLUTAMICUM

DECLARATION PURSUANT TO 37 C.F.R. § 1.132

I, Corey M. Crafton, declare as follows:

1. I have personal knowledge of the information contained herein.

2. I have over 10 years of experience with Archer-Daniels-Midland Company, including 6 years as a molecular biologist. My technical focus is on bacteria. I am also a registered patent agent.

3. I am a co-inventor of the subject matter claimed in U.S. Patent Application No. 09/978,763 ("the '763 application"), and as such I am familiar with the subject matter presented therein. I am also familiar with the prosecution of the '763 application. I have read and am familiar with the contents of the book excerpts and journal articles cited in this Declaration.

4. As one skilled in the art of molecular biology in general and bacterial engineering in particular, I recognize the utility of the invention described and claimed in the '763 application. I recognize that the invention as claimed has a specific and substantial utility, based at least on the factors discussed below.

5. One of ordinary skill in the art knows that a promoter is a nucleotide sequence that is recognized by RNA polymerase molecules which start RNA synthesis and that it is located immediately upstream of a gene. As explained in more detail in Devlin, T., *Textbook of Biochemistry with Clinical Correlations,* 689-696 (1997), a promoter consists of two highly conserved sequences: the -10 sequence (Pribnow box) and the -35 sequence. As stated in

Freifelder, D., *Molecular Biology: A Comprehensive Introduction to Prokaryotes and Eukaryotes*, 375-379 (1983), page 377, "All sequences found in Pribnow boxes are considered to be variants of the basic sequence TATAATG. The underscored T, at base 6 in the Pribnow box ... is present in all promoters sequenced to date." Figure 16.11 from Devlin, *supra,* page 690 shows these conserved sequences in many known E.Coli promoters.

When the Pribnow Box of SEQ ID NO 7 of the present invention is aligned into Figure 16.11 of Devlin, it is noted that only base 3 (C) is different from the most generally conserved sequence which has a T in the base 3 location. The most active promoters fit the consensus sequence most closely. The bases flanking the -10 and -35 sequences are only weakly conserved. Thus, the skilled person would ordinarily expect SEQ ID NO 7 to function as a promoter.

My project was to isolate several promoter regions from the Corynebacteria glutamicum lysine-producing strain. From research that had been done on promoters in E.Coli, a list of known E. Coli promoter sequences was assembled. The promoter upstream of the lactate dehydrogenase gene, ldh, was one these. From the professionally annotated complete genome sequence of Corynebacteria glutamicum, I located the genetic sequence that been annotated as the ldh gene. This annotation had been done by a professional organization that compared the Corynebacteria genome with publicly known and available genetic sequences from other organisms. The area of the Coryne genome that had the highest sequence identity to the known E. Coli ldh genetic sequence was therefore annotated as the Coryne ldh genetic sequence. At the time of this invention, the ldh promoter from Corynebacteria glutamicum had not been identified or annotated. I designed PCR primers to isolate a 500 bp fragment upstream of the annotated ldh gene since the ldh promoter should be upstream of the ldh coding region and should be between 20-200 bp long. As stated in Freifelder *supra*, page 375, "The first step in transcription is binding RNA polymerase to a DNA molecule. Binding occurs at particular sites called promoters, which are specific sequences of 20-200 bases at which several interactions occur." However, as one skilled in the art of molecular biology knows, 20-200 bp pieces of DNA are somewhat difficult to work with because of their small size. In order to make isolation and cloning steps easier, I designed the PCR primers to amplify a 500 bp piece. A 500 bp piece is large enough to ensure definite capture of the entire promoter region and an easy isolation from

an electrophoresis gel. As shown in the specification, a 500 bp PCR product was amplified for all potential promoter regions.

After isolation, the 500 bp piece was cloned into a screening vector to test for promoter activity.

6. Promoter utility is also shown by the β-galactosidase activity discussed in Example 9 of the '763 application: "Increased expression of beta-galactosidase under the transcriptional control of these transcriptional regulatory regions is shown in Table 9." (Paragraph [0202]). Based on my knowledge as one skilled in the art who has reviewed the data presented in Table 9 and throughout the specification, I would recognize that this increased activity is indicative of promoter activity because increased β-galactosidase activity is a conventionally used indicator of promoter activity in bacteria and fungi. Use of β-galactosidase activity as an indicator of promoter activity is discussed in, for example, Scanlan, D.J., *et al.*, "Construction of *lacZ* promoter probe vectors for use in *Synechococcus*: application to the identification of CO_2-regulated promoters," *Gene*, 90 (1990) 43-49; and Meyers, A.M., *et al.*, "Yeast shuttle and integrative vectors with multiple cloning sites suitable for construction of *lacZ* fusions," *Gene*, 45 (1986) 299-310.

7. Based on its sequence and on the functional data in Table 2, the regulator presented in SEQ ID NO: 7 includes the nucleotide sequence TACAATG in the -10 position (the "Pribnow Box") relative to the nucleotide sequence TTGCCAGGC in the -35 position. The Pribnow box in SEQ ID NO: 7 varies from the standard Pribnow box by only a single nucleotide (C instead of T at base 3), and includes the definitive T nucleotide at the base six position. When this element is positioned upstream of beta-galactosidase, the expression thereof is proof of promoter function and hence utility.

8. In fact, in a more recent sequence search using Genbank (http://www.ncbi.nlm.gov/blast), five highly conserved sequences were found. All five were Corynebacteria glutamicum sequences. One of these sequences (AB191244) is publicly annotated as the ldh promoter region. This sequence was submitted to Genbank on Mar 29, 2005. At the time of this invention, this sequence was not known or publicly available.

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like, so made, are punishable by fine or imprisonment, or both, under § 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of any patents issuing from the present application.

Date: Mar 7, 2006

By Corey M Crafton

Corey M. Crafton

Exhibit A

BLASTN 2.2.13 [Nov-27-2005]
Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. SchÃ¤ffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

RID: 1141056543-3021-6169699787.BLASTQ1

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
3,742,891 sequences; 16,670,205,594 total letters
Query=
Length=500

Sequences producing significant alignments:		Score (Bits)	E Value
gi|41326831|emb|BX927156.1|	Corynebacterium glutamicum ATCC 1...	944	0.0
gi|42602314|dbj|BA000036.3|	Corynebacterium glutamicum ATCC 1303	944	0.0
gi|80973081|gb|DQ248874.1|	Corynebacterium glutamicum L-lacta...	944	0.0
gi|62086196|dbj|AB191244.1|	Corynebacterium glutamicum ldhA gene	658	0.0
gi|50428370|dbj|AB115088.1|	Corynebacterium glutamicum ldhA g...	383	4e-103

ALIGNMENTS
>gi|41326831|emb|BX927156.1| Corynebacterium glutamicum ATCC 13032, IS fingerprint type 4-5,
complete genome; segment 9/10
Length=349115

Features in this part of subject sequence:
putative membrane protein

Score = 944 bits (476), Expect = 0.0
Identities = 494/500 (98%), Gaps = 0/500 (0%)
Strand=Plus/Minus

```
Query  1       AAAACAGCCAGGTTAGCGGCTGTAACCCACCACGGTTTCGGCAACAATGACGGCGAGAGA  60
               |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Sbjct  293611  AAAACAGCCAGGTTAGCAGCCGTAACCCACCACGGTTTCGGCAACAATGACGGCGAGAGA  293552

Query  61      GCCCACCACATTGCGATTTCCGCTCCGATAAAGCCAGCGCCCATATTTGCAGGGAGGATT  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  293551  GCCCACCACATTGCGATTTCCGCTCCGATAAAGCCAGCGCCCATATTTGCAGGGAGGATT  293492

Query  121     CGCCTGCGGTTTGGCGACATTCGGATCCCCGGAACCAGCTCTGCAATGACCTGCGCGCCG  180
               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  293491  CGCCTGCGGTTTGGCGACATTCGGATCCCCGGAACTAGCTCTGCAATGACCTGCGCGCCG  293432

Query  181     AGGGAAGCGAGGTGGGTGGCAGGTTTTAGTGCGGGTTTAAGCGTTGCCAGGCGAGTGGTG  240
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  293431  AGGGAGGCGAGGTGGGTGGCAGGTTTTAGTGCGGGTTTAAGCGTTGCCAGGCGAGTGGTG  293372
```

```
Query  241      AGCAAAGACGCTAGTCTGGGGAGCGAAACCATATTGAGTCATCTTGGCAGAGCATGCACA  300
                ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  293371   AGCAGAGACGCTAGTCTGGGGAGCGAAACCATATTGAGTCATCTTGGCAGAGCATGCACA  293312

Query  301      ATTCTGCAGGGCATAGATTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATA  360
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  293311   ATTCTGCAGGGCATAGGTTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATA  293252

Query  361      ACACTTGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGA  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  293251   ACACTTGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGA  293192

Query  421      ATCCGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAA  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  293191   ATCCGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAA  293132

Query  481      ATCAGGAAGTGGGATCGAAA  500
                ||||||||||||||||||||
Sbjct  293131   ATCAGGAAGTGGGATCGAAA  293112
```

>gi|42602314|dbj|BA000036.3| Corynebacterium glutamicum ATCC 13032 DNA, complete genome
Length=3309401

Features in this part of subject sequence:
  Hypothetical protein

Score = 944 bits (476), Expect = 0.0
Identities = 494/500 (98%), Gaps = 0/500 (0%)
Strand=Plus/Minus

```
Query  1        AAAACAGCCAGGTTAGCGGCTGTAACCCACCACGGTTTCGGCAACAATGACGGCGAGAGA  60
                |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Sbjct  3113891  AAAACAGCCAGGTTAGCAGCCGTAACCCACCACGGTTTCGGCAACAATGACGGCGAGAGA  3113832

Query  61       GCCCACCACATTGCGATTTCCGCTCCGATAAAGCCAGCGCCCATATTTGCAGGGAGGATT  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3113831  GCCCACCACATTGCGATTTCCGCTCCGATAAAGCCAGCGCCCATATTTGCAGGGAGGATT  3113772

Query  121      CGCCTGCGGTTTGGCGACATTCGGATCCCCGGAACCAGCTCTGCAATGACCTGCGCGCCG  180
                |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  3113771  CGCCTGCGGTTTGGCGACATTCGGATCCCCGGAACTAGCTCTGCAATGACCTGCGCGCCG  3113712

Query  181      AGGGAAGCGAGGTGGGTGGCAGGTTTTAGTGCGGGTTTAAGCGTTGCCAGGCGAGTGGTG  240
                |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3113711  AGGGAGGCGAGGTGGGTGGCAGGTTTTAGTGCGGGTTTAAGCGTTGCCAGGCGAGTGGTG  3113652

Query  241      AGCAAAGACGCTAGTCTGGGGAGCGAAACCATATTGAGTCATCTTGGCAGAGCATGCACA  300
                ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3113651  AGCAGAGACGCTAGTCTGGGGAGCGAAACCATATTGAGTCATCTTGGCAGAGCATGCACA  3113592

Query  301      ATTCTGCAGGGCATAGATTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATA  360
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  3113591  ATTCTGCAGGGCATAGGTTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATA  3113532
```

```
Query  361      ACACTTGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGA  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3113531  ACACTTGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGA  3113472

Query  421      ATCCGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAA  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3113471  ATCCGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAA  3113412

Query  481      ATCAGGAAGTGGGATCGAAA  500
                ||||||||||||||||||||
Sbjct  3113411  ATCAGGAAGTGGGATCGAAA  3113392
```

>gi|80973081|gb|DQ248874.1| Corynebacterium glutamicum L-lactate dehydrogenase (ldh) and
pyruvate kinase (pyk) genes, complete cds
Length=4183

```
 Score =  944 bits (476),  Expect = 0.0
 Identities = 494/500 (98%), Gaps = 0/500 (0%)
 Strand=Plus/Plus

Query  1    AAAACAGCCAGGTTAGCGGCTGTAACCCACCACGGTTTCGGCAACAATGACGGCGAGAGA  60
            |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Sbjct  103  AAAACAGCCAGGTTAGCAGCCGTAACCCACCACGGTTTCGGCAACAATGACGGCGAGAGA  162

Query  61   GCCCACCACATTGCGATTTCCGCTCCGATAAAGCCAGCGCCCATATTTGCAGGGAGGATT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  163  GCCCACCACATTGCGATTTCCGCTCCGATAAAGCCAGCGCCCATATTTGCAGGGAGGATT  222

Query  121  CGCCTGCGGTTTGGCGACATTCGGATCCCCGGAACCAGCTCTGCAATGACCTGCGCGCCG  180
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  223  CGCCTGCGGTTTGGCGACATTCGGATCCCCGGAACTAGCTCTGCAATGACCTGCGCGCCG  282

Query  181  AGGGAAGCGAGGTGGGTGGCAGGTTTTAGTGCGGGTTTAAGCGTTGCCAGGCGAGTGGTG  240
            ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  283  AGGGAGGCGAGGTGGGTGGCAGGTTTTAGTGCGGGTTTAAGCGTTGCCAGGCGAGTGGTG  342

Query  241  AGCAAAGACGCTAGTCTGGGGAGCGAAACCATATTGAGTCATCTTGGCAGAGCATGCACA  300
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  343  AGCAGAGACGCTAGTCTGGGGAGCGAAACCATATTGAGTCATCTTGGCAGAGCATGCACA  402

Query  301  ATTCTGCAGGGCATAGATTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATA  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  403  ATTCTGCAGGGCATAGATTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATA  462

Query  361  ACACTTGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGA  420
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  463  ACACATGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGA  522

Query  421  ATCCGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  523  ATCCGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAA  582

Query  481  ATCAGGAAGTGGGATCGAAA  500
            ||||||||||||||||||||
Sbjct  583  ATCAGGAAGTGGGATCGAAA  602
```

>gi|62086196|dbj|AB191244.1| Corynebacterium glutamicum ldhA gene, promoter region
Length=348

 Score = 658 bits (332), Expect = 0.0
 Identities = 344/348 (98%), Gaps = 0/348 (0%)
 Strand=Plus/Plus

```
Query  150  CGGAACCAGCTCTGCAATGACCTGCGCGCCGAGGGAAGCGAGGTGGGTGGCAGGTTTTAG  209
            |||||| |||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  1    CGGAACTAGCTCTGCAATGACCTGCGCGCCGAGGGAGGCGAGGTGGGTGGCAGGTTTTAG  60

Query  210  TGCGGGTTTAAGCGTTGCCAGGCGAGTGGTGAGCAAAGACGCTAGTCTGGGGAGCGAAAC  269
            |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  61   TGCGGGTTTAAGCGTTGCCAGGCGAGTGGTGAGCAGAGACGCTAGTCTGGGGAGCGAAAC  120

Query  270  CATATTGAGTCATCTTGGCAGAGCATGCACAATTCTGCAGGGCATAGATTGGTTTTGCTC  329
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CATATTGAGTCATCTTGGCAGAGCATGCACAATTCTGCAGGGCATAGATTGGTTTTGCTC  180

Query  330  GATTTACAATGTGATTTTTTCAACAAAAATAACACTTGGTCTGACCACATTTTCGGACAT  389
            |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  181  GATTTACAATGTGATTTTTTCAACAAAAATAACACATGGTCTGACCACATTTTCGGACAT  240

Query  390  AATCGGGCATAATTAAAGGTGTAACAAAGGAATCCGGGCACAAGCTCTTGCTGATTTTCT  449
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  AATCGGGCATAATTAAAGGTGTAACAAAGGAATCCGGGCACAAGCTCTTGCTGATTTTCT  300

Query  450  GAGCTGCTTTGTGGGTTGTCCGGTTAGGGAAATCAGGAAGTGGGATCG  497
            |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  GAGCTGCTTTGTGGGTTGTCCGGTTAGGGAAATCAGGAAGTGGGATCG  348
```

>gi|50428370|dbj|AB115088.1| Corynebacterium glutamicum ldhA gene for lactate dehydrogenase,
complete cds
Length=1456

 Score = 383 bits (193), Expect = 4e-103
 Identities = 196/197 (99%), Gaps = 0/197 (0%)
 Strand=Plus/Plus

```
Query  304  CTGCAGGGCATAGATTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATAACA  363
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    CTGCAGGGCATAGATTGGTTTTGCTCGATTTACAATGTGATTTTTTCAACAAAAATAACA  60

Query  364  CTTGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGAATC  423
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CATGGTCTGACCACATTTTCGGACATAATCGGGCATAATTAAAGGTGTAACAAAGGAATC  120

Query  424  CGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAAATC  483
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CGGGCACAAGCTCTTGCTGATTTTCTGAGCTGCTTTGTGGGTTGTCCGGTTAGGGAAATC  180

Query  484  AGGAAGTGGGATCGAAA  500
            |||||||||||||||||
```

Sbjct 181 AGGAAGTGGGATCGAAA 197

Gene, 90 (1990) 43–49
Elsevier

GENE 03575

Construction of *lacZ* promoter probe vectors for use in *Synechococcus*: application to the identification of CO_2-regulated promoters

(Gene fusions; β-galactosidase; inorganic carbon uptake; recombinant DNA; promoter; plasmid)

David J. Scanlan, Stephen A. Bloye, Nicholas H. Mann, David A. Hodgson and Noel G. Carr

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL (U.K.)

Received by K.F. Chater: 7 November 1989
Revised: 2 January 1990
Accepted: 4 January 1990

SUMMARY

It was shown that the *Escherichia coli lacZ* gene could be expressed in the cyanobacterium *Synechococcus* R2 PCC7942 both as a plasmid-borne form and also integrated into the chromosome. A promoterless form of the *lacZ* gene was constructed and used as a reporter gene to make transcriptional fusions with cyanobacterial promoters using a shuttle vector system and also via a process of integration by homologous recombination. *Synechococcus* R2 promoter-*lacZ* gene fusions were then used to identify CO_2-regulated promoters, by quantitatively assessing β-galactosidase activity under high and low CO_2 conditions using a fluorescence assay. Several promoters induced under low CO_2 conditions were detected.

INTRODUCTION

Cyanobacteria are capable of oxygen-evolving photosynthesis and the utilization of CO_2 as a sole source of carbon. Unicellular cyanobacteria offer attractive systems for the study of photosynthesis, particularly so when gene transfer is possible by either natural or recombinant means. *Synechococcus* R2 PCC7942 is an organism which is most readily transformed, and efficient shuttle vectors have been in use for some time (Kuhlemeier et al., 1981).

Correspondence to: Dr. D.J. Scanlan, Department of Biological Sciences, University of Warwick, Coventry CV4 7AL (U.K.) Tel. (0203)523544; Fax (0203)523701.

Abbreviations: aa, amino acid(s); Ap, ampicillin; βGal, β-galactosidase; bp, base pair(s); C_i, inorganic carbon; Cm, chloramphenicol; *cpc*, gene encoding phycocyanin; DMSO, dimethylsulfoxide; kb, kilobase(s) or 1000 bp; Km, kanamycin; K_m, Michaelis–Menten constant; MCS, multiple cloning site; MUG, 4-methyl umbelliferyl-β-D-galactopyranoside; nt, nucleotide(s); ONPG, *o*-nitrophenyl-β-D-galactopyranoside; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; ^{R}, resistant/resistance; RuBisCO, D-ribulose 1,5-bisphosphate carboxylase/oxygenase; TE, 10 mM Tris/1 mM EDTA pH 8.0; XGal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; [ ], denotes plasmid-carrier state.

Cyanobacteria, together with some unicellular eukaryotic phototrophs and certain lower aquatic plants, are capable of concentrating exogenous bicarbonate (see Badger, 1987) thereby providing a high internal concentration of CO_2 for RuBisCO and overcoming the high K_m of this enzyme for its substrate. Recently, mutants have been isolated which require high CO_2 conditions to grow (see Marcus et al., 1986; Abe et al., 1988) and their use may yield information on the molecular nature of C_i uptake.

Various reporter genes, e.g., *cat* (Friedberg and Seijffers, 1986) and *lux* (Schmetterer et al., 1986) have been used to assess the expression of specific genes in cyanobacteria. Expression of βGal in the marine cyanobacterium *Synechococcus* sp. PCC7002, has been reported (Buzby et al., 1985) and applied to assessment of the effect of light intensity and nitrogen availability on *cpc-lacZ* gene fusions (Gasparich et al., 1987). Such studies show that *lacZ* gene fusions can be used to monitor gene expression in cyanobacteria. It was consequently decided to take the approach that cyanobacterial DNA-*lacZ* gene fusions could be used to identify presumptive CO_2-regulated promoters by the differential activity of βGal under high and low CO_2 conditions.

RESULTS AND DISCUSSION

(a) Synthesis of βGal in *Synechococcus* R2-SPc PCC7942

Two strategies are applicable, in cyanobacteria, to the construction of recombinants in which the expression of a reporter gene is driven from a cyanobacterial promoter. Either the hybrid may be introduced into the chromosome by homologous recombination or be expressed from an independently replicating shuttle vector (Table I). Using the latter approach we have obtained gene fusions in which expression of the reporter gene was driven by promoters whose activity was controlled by CO_2 availability. It was decided to use *lacZ* as the reporter gene since it had been shown to be efficiently expressed in the unicellular cyanobacterium *Synechococcus* PCC7002 (Buzby et al., 1985).

To confirm that the *lacZ* gene could be expressed in *Synechococcus* R2-SPc it was introduced into the *E. coli*/*Synechococcus* shuttle vector pUC105 (Kuhlemeier et al., 1981), by ligating a 4.2-kb *Eco*RI-*Sal*I fragment from pTEBG3 into *Eco*RI + *Sal*I-digested pUC105. The resulting 14-kb plasmid pTUC1 was introduced into *Synechococcus* R2-SPc. Ap^R transformants were obtained at a frequency of 10^2–$10^3/\mu g$ DNA. The presence of the *lacZ* gene was confirmed by Southern blotting and by the fluorescence of transformants after spraying with MUG. (We found that XGal was a less suitable indicator because the endogenous pigmentation of *Synechococcus* colonies obscured the indicator colour.) In addition, MUG can only detect βGal present in bacteria at the time the substrate is applied, in contrast with colour reactions produced by bacterial colonies grown on agar plates that include XGal, which reflect substrate hydrolysis throughout the development of the colony.

(b) Construction of generalised promoter-probe vectors

Fig. 1A describes the construction of this new *lacZ* fragment and of pDAH216 and pDAH274. The sequence around the 5′ end of *lacZ* showing the fusion point with *trpA* and stop codons is illustrated in Fig. 1B.

(c) Construction of the *lacZ* promoter probes for use in *Synechococcus*

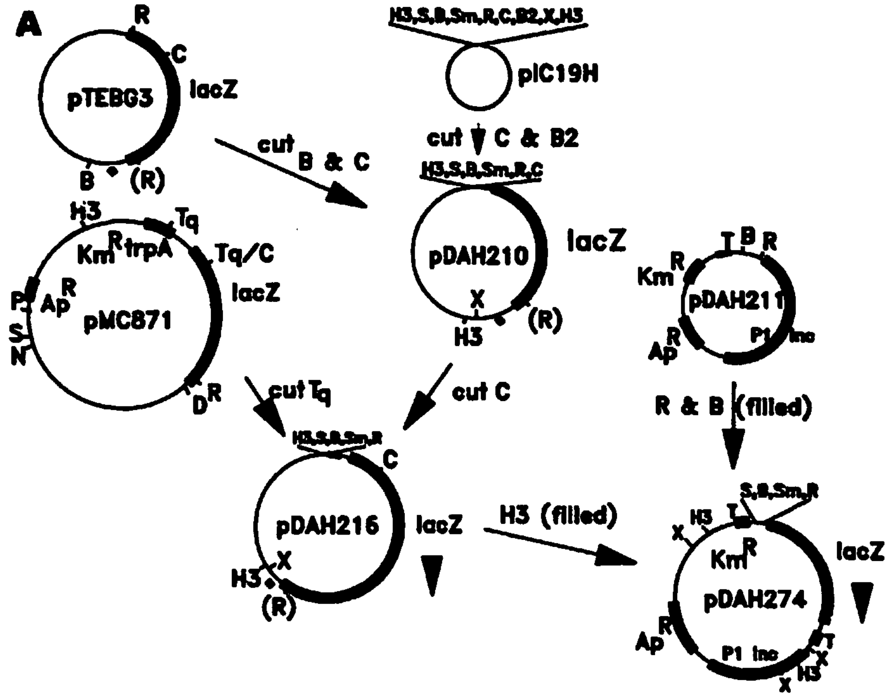
Fig. 2 describes construction of the new *lacZ* promoter probe plasmids, based on the shuttle vector pUC105

TABLE I

Bacterial strains and plasmids

Strain or plasmid[a]	Characteristics[a]	Source/reference
Escherichia coli		
DH1	F^-, *recA*1, *endA*1, *gyrA*96	Maniatis et al. (1982)
MC1061	*araD*139, Δ(*ara-leu*)7697, Δ*lac*X74, *galK*$^-$, *hsdR*$^-$, *hsdM*$^+$, *strA*	Casabadan and Cohen (1980)
Synechococcus		
PCC7942	R2-SPc (small plasmid cured)	This laboratory
Plasmids		
E. coli		
pTEBG3	Ap^R, pBR322::*lacZ*	S. Elledge
pREG422	Ap^R	Shimkets et al. (1983)
pIC19H	Ap^R	Marsh et al. (1984)
pDAH216	Ap^R, promoterless *lacZ*	This study
pDAH274	Ap^R, Km^R, promoterless *lacZ*, P1 *inc*	This study
Synechococcus		
pUC105	Ap^R, Cm^R	Kuhlemeier et al. (1981)
pUC303	Sm^R, Cm^R	Kuhlemeier et al. (1983)
pUC105XS	Ap^R, Cm^R, *Xho*I-*Sal*I deletion	This study
pUC105H	Cm^R, *Hin*dIII deletion of pUC105	This study
pTUC1	Cm^R, Ap^R *lacZ*	This study
pLACPB1	Ap^R, Cm^R, promoterless *lacZ*	This study
pLACPB2	Ap^R, Cm^R, transcription terminators, promoterless *lacZ*	This study

[a] The cyanobacterium *Synechococcus* R2-SPc PCC7942 was grown at 34°C in Allen's medium (Allen, 1968) in an orbital shaker and illuminated at a light intensity of 30–40 $\mu E/m^2/s$. Low CO_2 cultures were grown in an environment gassed with air and high CO_2 cultures in a gas phase of 5% (v/v) CO_2 in air with Allen's medium supplemented with 10 mM $NaHCO_3$. Solid medium contained 1.5% (w/v) Bacto agar with the agar and Allen's medium autoclaved separately. Antibiotic concentrations used for *Synechococcus* R2-SPc grown in liquid medium were 1 μg Ap/ml and 10 μg Cm/ml. *E. coli* was grown in nutrient broth or on nutrient agar at 37°C. Antibiotic concentrations in both liquid and solid medium were 50 μg Ap/ml and 30 μg Cm/ml. Δ, deletion; P1 inc, incompatability region for phage P1.



B

```
                    -35                        -10              operator                      RBS          Met Thr Met Ile
lacZ 5' Region AG GCTTTACACTTT ATGCTTCCGGCTCG TATGTT GTGTG TGGAATTGTGAGCGGATAACAATT T CACAC AGGA AACAGCT ATG ACC ATG ATT

                Ala Ala Thr Arg Ser End
trpA 3' Region  GCG GCG ACG CGC AGT TAA TCCCACAGCCGCCAGTTCCGCTGGCGGCATTTTAACTTTCTTTAATGGAAGCCGGAAAAATCCTAAA

                                                                                              RBS          Met Thr Met Ile
                Ala Ala Thr Arg Ser End
trpA/lacZ       GCG GCG ACG CGC AGT TAA TCCCACAGCCGCCAGTTCCGCTGGCGGCATTTTAACTTTCTTTAT CACAC AGGA AACAGCT ATG ACC ATG ATT
Fusion ΔW205
```

Fig. 1. Construction of the generalised *lacZ* promoter probe plasmids pDAH216 and pDAH274. Plasmids are not drawn to scale. The promoterless *lacZ* gene of pMC871 (Casadaban et al., 1980) was modified to: (*i*) remove as much as possible of the unwanted *trpA* DNA upstream from the *lacZ*, whilst still retaining stop codons in all three reading frames, thus ensuring that protein fusions were not possible; (*ii*) remove the *Eco*RI restriction site in the C-terminal region of the gene without changing the aa sequence; and (*iii*) insert a MCS upstream from the *lacZ* gene for the insertion of potential promoter fragments. The MCS includes a, now unique, *Eco*RI restriction site. Plasmid pTEBG3 contains a modified *lacZ* gene in which the *Eco*RI site was destroyed by site-directed mutagenesis. However, the same aa are still encoded in the region of the modification. (A) Construction of the modified *lacZ* fragment and of pDAH216 and pDAH274. A 2.6-kb *Bam*HI-*Cla*I fragment of pTEBG3 containing the C-terminal half of *lacZ* was ligated to *Bgl*II + *Cla*I-cut pIC19H, to produce pDAH210. *Cla*I-cut pDAH210 was then ligated to *Taq*I-digested pMC871. The resulting plasmid, pDAH216, contains a shortened *lacZ* fragment lacking the *Eco*RI site in the C-terminal end of the *lacZ* gene. There are many *Taq*I sites present in pMC871 but none within the *Taq*I fragment containing the N-terminal end of the *lacZ* gene indicated. The other *Taq*I sites have been omitted for clarity. *Hin*dIII cut pDAH216 (and ends filled-in) was then ligated to *Eco*RI + *Bam*HI-cut (and ends filled-in) pDAH211 (derived from pMC871 and pREG422; see Shimkets et al., 1983) to produce pDAH274. Abbreviations: B, *Bam*HI; B2, *Bgl*II; C, *Cla*I; D, *Dra*I; H3, *Hin*dIII; K, *Kpn*I; N, *Nco*I; P, *Pst*I; R, *Eco*RI; (R), deleted *Eco*RI site; S, *Sal*I; Sm, *Sma*I; Tq, *Taq*I; X, *Xho*I; T, transcription termination signal; ◆, terminator from *Bacillus subtilis* α-amylase gene. Thin lines represent plasmid sequences, whilst thick lines represent drug resistance genes, *lacZ*, P1 inc (incompatibility region for phage P1) or transcription termination signals. (B) Sequence of the 5′ end of *lacZ* and the 3′ end of *trpA* (Casadaban et al., 1980) showing the *trpA*/*lacZ* fusion point of pDAH216 and pDAH274 (S. McGowan, personal communication). Italicized letters represent bp not present in the *trpA*/*lacZ* fusion of ΔW205. RBS, ribosome-binding site.

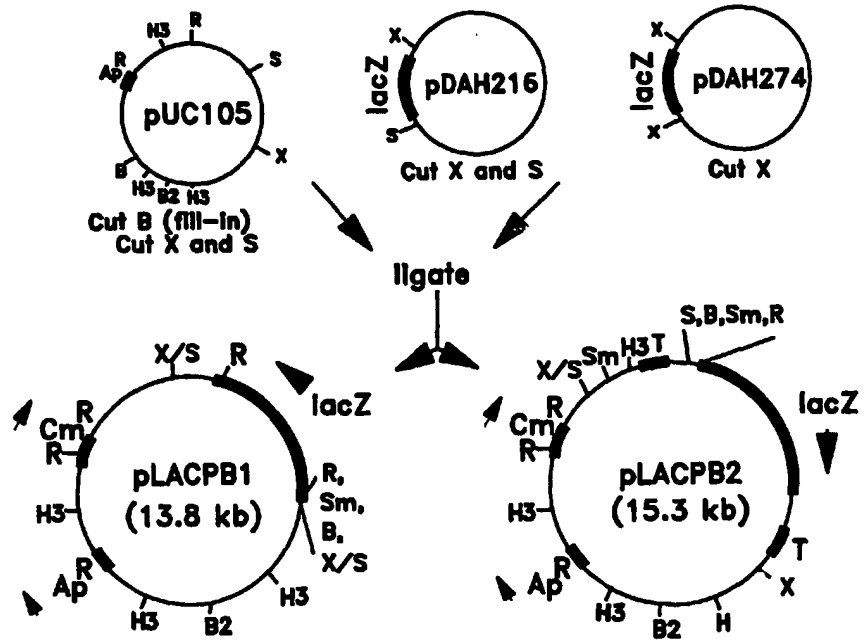


Fig. 2. Construction of *Synechococcus lacZ* promoter probes. The *Bam*HI site in pUC105 was removed by filling-in cut ends using PolIk, and the resulting vector (pUC105 Bam-) cut with *Xho*I + *Sal*I. Deletion experiments with pUC105 showed that a *Xho*I-*Sal*I deletion did not affect transformation frequency (Table II). This was in contrast to deletion of a 4-kb *Hin*dIII fragment from pUC105 which completely abolished replication of pUC105 in *Synechococcus* R2. This would agree with the proposal that the cyanobacterial replication origin of this plasmid is contained within a 4.65-kb *Bam*HI-*Xho*I restriction fragment (Gendel, 1987). Insertion of a promoterless *lacZ* gene from pDAH216, or pDAH274 (containing transcription termination signals at either end of the *lacZ* gene), into the large *Xho*I-*Sal*I fragment of pUC105 Bam- produced pLACPB1 and pLACPB2, respectively.

(Kuhlemeier et al., 1981) for use in *Synechococcus* R2 PCC7942. Both vectors contain a unique *Bam*HI site for insertion of cyanobacterial chromosomal DNA. *Synechococcus* R2-SPc chromosomal DNA libraries were constructed with these vectors, cloning partial *Sau*3AI digested chromosomal DNA into this unique *Bam*HI site. Transformation frequencies of up to 10^6 transformants/μg DNA were obtained (Table II).

Although plasmid promoter probe vectors have been widely used for studying gene fusions, the background expression of an intact *lacZ* gene on a multicopy plasmid, even lacking a recognisable promoter, is apparently quite high (Casadaban et al., 1980). Thus, it may be difficult to distinguish strains carrying the desired fusion from those carrying the parent plasmid. pLACPB1 indeed exhibited some endogenous βGal activity (Table III). However, using pLACPB2 which contained a transcription termination signal upstream from the site of insertion of chromosomal DNA, expression of βGal was reduced twofold. This difference reflects a transcriptional effect since the plasmid origin of replication (and hence the plasmid copy number) is the same in each case.

(d) Use of pLACPB1 and pLACPB2 to identify CO_2-regulated promoters

Synechococcus R2-SPc chromosomal DNA libraries, constructed in pLACPB1 and pLACPB2 with approx. 4-kb fragments generated by *Sau*3AI partial digestion, were used to transform *Synechococcus* R2-SPc under normal low CO_2 conditions. Transformants were restreaked onto Allen's medium containing 7.5 μg Cm/ml, and then replica-plated onto solid medium containing Cm + 10 mM $NaHCO_3$. Plates were placed inside sealed gas bags before gassing with 5% CO_2 in air. After five days, corresponding high and low CO_2 transformation plates were sprayed with MUG and photographed. This initial screening allowed a preliminary identification of transformants exhibiting CO_2-regulated expression of βGal (Fig. 3). Interesting transformants were then grown in liquid medium under high and low CO_2 conditions, and βGal was assayed throughout the growth curve using the MUG assay. Generally, βGal activity increased proportionately with growth, though a few transformants showed a slight decrease when reaching stationary phase. Differences in βGal activity under high or low CO_2 conditions were observed in 8 of 600 pLACPB1 or 17 of 2500 pLACPB2 transformants screened – showing either greater or lesser βGal activity under the different CO_2 concentrations. Table III shows some examples. CO_2 concentration did not significantly affect βGal activity in control cultures.

Recent observations suggest that light intensity may also have a controlling effect on *lacZ* expression of individual transformants (data not shown). This is in agreement with the idea that metabolic conditions within the cell might be

TABLE II

Transformation frequencies for *Synechococcus* R2-SPc shuttle vector, integrative vector and promoter-probe vectors

Plasmid/selection[a]		Transformants/ μg DNA[b]
pUC105	Cm^R	10^5
pUC105 *Xho*I-*Sal*I deletion	Cm^R	10^4
pUC105 *Hin*dIII deletion	Cm^R	zero
pLACPB1	Cm^R	10^6
pLACPB1 chromosomal DNA library	Cm^R	10^5–10^6
pLACPB2	Cm^R	10^4
pLACPB2 chromosomal DNA library	Cm^R	10^2–10^3
pTUC1	Ap^R	10^2–10^3
pDAH274	Km^R	zero
pDAH274 chromosomal DNA library 4-kb DNA fragments	Km^R	10^3–10^4

[a] Small-scale plasmid isolation from cyanobacteria used the rapid boiling method of Holmes and Quigley (1981) as modified by Alley (1987). Cyanobacterial chromosomal DNA extraction was based on a method described by Lind et al. (1985) with modifications. A late-log phase culture (25 ml) was spun in a MSE multex centrifuge at 5000 rpm for 10 min, resuspended in 0.5 ml 0.25 M Tris pH 8.0/20% (w/v) sucrose/lysozyme 10 mg per ml, and the cells were incubated for 1 h at 37°C. Sarkosyl (16 μl of 30% (v/v) solution) and 20 μl of proteinase K (5 mg/ml) was then added, and the cells incubated at 65°C for 1 h. An equal volume of phenol : chloroform was added, and the mixture vortexed and spun for 4 min in an eppendorf centrifuge. The supernatant was dialysed overnight against TE buffer and stored at –20°C. Plasmid constructions and transformation of *E. coli* were performed by standard techniques described in Maniatis et al. (1982). Restriction enzymes (Amersham International) were used under conditions recommended by the manufacturers.

[b] Transformation of *Synechococcus* R2-SPc PCC7942 was performed as described by Kuhlemeier et al. (1981). Where appropriate, transformants were replica plated onto solid medium containing 10 mM $NaHCO_3$ plus antibiotic (7.5 μg Cm/ml or 1 μg Ap/ml). These plates were placed inside sealed plastic bags containing an atmosphere of 5% (v/v) CO_2 in air and continuously illuminated.

similar under low CO_2 levels and high light intensities, and follows the identification of a 42-kDa cytoplasmic membrane protein from *Synechococcus* R2 which has been shown to be regulated by CO_2 concentration and light intensity (Reddy et al., 1989). Using both MUG and ONPG assays it was shown that many of the CO_2-regulated promoters were functional in *E. coli* (data not shown).

We have recently constructed a *Synechococcus* R2 gene library directly into pDAH274, a vector incapable of independent replication in cyanobacteria. Using this insert-directed integration system transformants were obtained at high frequency (Table II), were stable in the presence of Km, and showed differential *lacZ* expression. pDAH274 without inserts failed to transform *Synechococcus* R2-SPc. The control of the *lacZ* gene from cyanobacterial promoters maintained solely on the chromosome simplifies problems of plasmid copy number.

TABLE III

Expression of *lacZ* in selected *Synechococcus* R2-SPc transformants grown under low and high CO_2 conditions

Transformant[a]	βGal activity (MUG units[b])		Ratio of βGal activity (low CO_2/high CO_2)
	Air level CO_2	5% (v/v) CO_2 in air	
Synechococcus R2-SPc (untransformed)	0.6	0.6	1.0
pLACPB1 control	46.0	32.0	1.4
8	29.0	117.0	0.25
10	1560.0	450.0	3.5
14	82.0	29.0	2.8
19	2630.0	79.0	33.0
pLACPB2 control	24.0	17.0	1.4
4	925.0	232.0	4.0
5	1955.0	223.0	9.0
9	370.0	57.0	6.5
17	423.0	42.0	10.0

[a] See Table I. Transformations were carried out as described in Table II, footnote b. Nos. 8, 10, 14 and 19 represent specific pLACPB1 chromosomal DNA library transformants. Nos. 4, 5, 9 and 17 are specific pLACPB2 chromosomal DNA library transformants.

[b] βGal activity was assayed using either ONPG as described by Miller (1972) and data are expressed as the increase in A_{420}/min/ml/mg protein, or using MUG, a quantitative fluorimetric assay for βGal specific activity, carried out as described by Youngman (1987). MUG units represent pmol MUG hydrolysed/ml/min standardised for culture density.

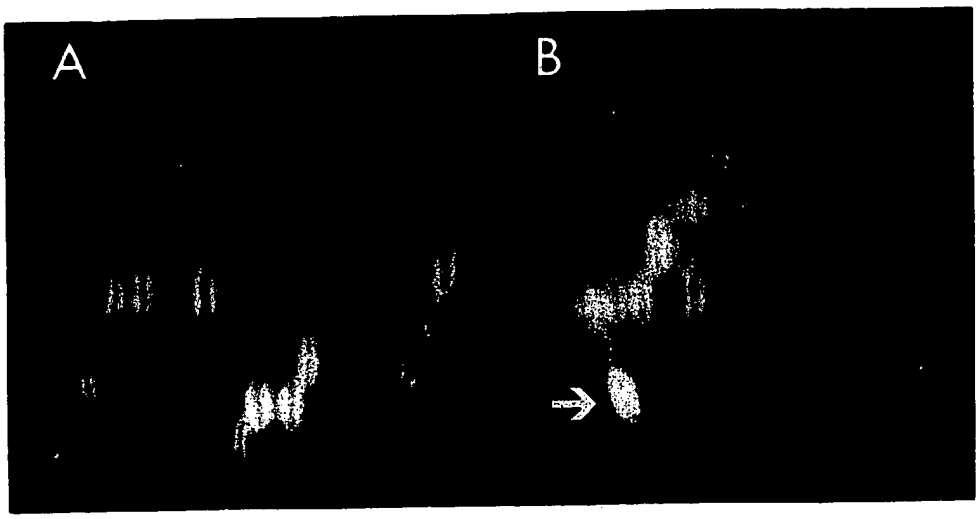


Fig. 3. Differential *lacZ* activity by *Synechococcus* R2-SPc-[pLACPB2] transformants grown under high (A) and low (B) CO_2 conditions as assessed by spraying plates with MUG. MUG was applied after patched bacterial colonies had developed, by spraying the plate with a MUG solution (10 mg/ml in DMSO). Plates were held 30 cm away from the atomizer nozzle, and a fine spray of MUG was delivered over the surface of the plate. After 5–10 min, plates were visualised under long wavelength ultraviolet light and photographed using Polaroid 667 film at f11 for an 1/8 of a second using a Kodak No. 45 Wratten gelatin filter. Magnification, × 0.7. The arrow indicates a transformant showing greater βGal activity under low CO_2 conditions.

(e) Conclusions

This study describes the construction of *lacZ* promoter probe vectors and their modification and use in the unicellular cyanobacterium *Synechococcus* R2 PCC7942.

(1) The *lacZ* gene was shown to be expressed in this organism from both an endogenously replicating plasmid and also integrated into the chromosome.

(2) Plasmids pLACPB1 and pLACPB2 are *lacZ* promoter probes for use in *Synechococcus* R2 with a replication origin functional in this organism, and which transform *Synechococcus* R2 at high frequency. In addition to the various presumptive CO_2-regulated promoters described here we have also identified promoters regulated by iron and magnesium limitation (data not shown). These plasmids allow relatively easy isolation of the promoter fragment which can then be used to clone the whole gene which would enable various functional studies. This approach thus allows an alternative molecular approach to studying for example inorganic carbon uptake in this organism.

ACKNOWLEDGEMENTS

This work was supported in part by a SERC-CASE studentship in association with Celltech to D.J.S. and a SERC studentship to S.A.B. and by the AFRC. D.A.H. acknowledges the support of a 'Nuffield Foundation Award to Newly Appointed Lecturers in Science' for part of this work.

REFERENCES

Abe, T., Tsuzuki, M., Kadokami, Y. and Miyachi, S.: Isolation and characterisation of temperature-sensitive, high-CO_2 requiring mutant of *Anacystis nidulans* R2. Plant Cell Physiol. 29 (1988) 1353–1360.

Allen, M.M.: Simple conditions for growth of unicellular blue-green algae on plates. J. Phycol. 4 (1968) 1–4.

Alley, M.R.K.: Molecular Biological Aspects of Nitrogen Starvation in Cyanobacteria. Ph.D. Thesis, University of Warwick, 1987.

Badger, M.R.: The CO_2 concentrating mechanism in aquatic prototrophs. In Hatch, M.D. and Broudman, N.K. (Eds.), The Biochemistry of Plants: A Comprehensive Treatise, Vol. 10, Photosynthesis. Academic Press, San Diego, 1987, pp. 219–274.

Buzby, J.S., Porter, R.D. and Stevens Jr., S.E.: Expression of the *Escherichia coli lacZ* gene on a plasmid vector in a cyanobacterium. Science 230 (1985) 805–807.

Casadaban, M.J. and Cohen, S.N.: Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138 (1980) 179–207.

Casadaban, M.J., Chou, J. and Cohen, S.N.: In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. 143 (1980) 971–980.

Friedberg, D.: Use of reporter genes in cyanobacteria. Methods Enzymol. 167 (1988) 736–747.

Friedberg, D. and Seijffers, J.: Controlled gene expression utilising lambda phage regulatory signals in a cyanobacterium host. Mol. Gen. Genet. 203 (1986) 505–510.

Gasparich, G., Buzby, J.S., Bryant, D.A., Porter, R.D. and Stevens Jr., S.E.: The effects of light intensity and nitrogen starvation on the phycocyanin promoter in the cyanobacterium *Synechococcus* PCC7002. In Biggins, J. (Ed.), Progress in Photosynthetic Research, Vol. 4. Nijhoff, Dordrecht, 1987, pp. 761–765.

Gendel, S.M.: Instability of Tn*5* inserts in cyanobacterial cloning vectors. J. Bacteriol. 169 (1987) 4426–4430.

Holmes, D.S. and Quigley, M.: A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114 (1981) 193.

Kuhlemeier, C.J., Borrias, W.E., Van den Hondel, C.A.M.J.J. and Van Arkel, G.A.: Construction and characterisation of two hybrid plasmids capable of transformation to *Anacystis nidulans* R2 and *Escherichia coli* K-12. Mol. Gen. Genet. 184 (1981) 249–254.

Kuhlemeier, C.J., Thomas, A.A.M., Van der Ende, A., Van Leen, R.W., Borrias, W.E., Van den Hondel, C.A.M.J.J. and Van Arkel, G.A.: A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. Plasmid 10 (1983) 156–163.

Lind, L.K., Kalla, S.R., Lonneborg, A., Oquist, G. and Gustafsson, P.: Cloning of the β-phycocyanin gene from *Anacystis nidulans*. FEBS Lett. 188 (1985) 27–32.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Marcus, Y., Schwarz, R., Friedberg, D. and Kaplan, A.: High CO_2 requiring mutant of *Anacystis nidulans* R2. Plant Physiol. 82 (1986) 610–612.

Marsh, J.L., Erfle, M. and Wykes, E.J.: The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene 32 (1984) 481–485.

Miller, J.H.: Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1972.

Reddy, K.J., Masamoto, K., Sherman, D.M. and Sherman, L.A.: DNA sequence and regulation of the gene (*cbpA*) encoding the 42-kilodalton cytoplasmic membrane carotenoprotein of the cyanobacterium *Synechococcus* sp. strain PCC 7942. J. Bacteriol. 171 (1989) 3486–3493.

Schmetterer, G., Wolk, C.P. and Elhai, J.: Expression of luciferases from *Vibrio harveyi* and *Vibrio fischeri* in filamentous cyanobacteria. J. Bacteriol. 167 (1986) 411–414.

Shimkets, L.J., Gill, R.E. and Kaiser, D.: Developmental cell interactions in *Myxococcus xanthus* and the *spoC* locus. Proc. Natl. Acad. Sci. USA 80 (1983) 1406–1410.

Youngman, P.: Plasmid vectors for recovering and exploiting Tn*917* transpositions in Bacillus and other Gram-positive bacteria. In Hardy, K.G. (Ed.), Plasmids. A Practical Approach. IRL Press, Oxford, 1987, pp. 93–103.

Gene, 45 (1986) 299–310
Elsevier

GENE 1716

Yeast shuttle and integrative vectors with multiple cloning sites suitable for construction of *lacZ* fusions

(Recombinant DNA; plasmid; promoter; yeast; *Escherichia coli*; codon; reading frame; catabolite repression; β-galactosidase)

Alan M. Myers[a], Alexander Tzagoloff[a], Dennis M. Kinney[b]* and Carol J. Lusty[b]

[a] Department of Biological Sciences, Columbia University, New York, NY 10027, Tel. (212)280-2920, and [b] Public Health Research Institute of the City of New York, Inc., 455 First Ave., New York, NY 10016 (U.S.A.) Tel. (212)578-0827

(Received April 7th, 1986)
(Accepted June 3rd, 1986)

SUMMARY

We report yeast/*Escherichia coli* shuttle vectors suitable for fusing yeast promoter and coding sequences to the *lacZ* gene of *E. coli*. The vectors contain a region of multiple unique restriction sites including *Eco*RI, *Kpn*I, *Sma*I, *Bam*HI, *Xba*I, *Sal*I, *Pst*I, *Sph*I and *Hin*dIII. The region with the unique cloning sites has been introduced in both orientations with respect to *lacZ* and occurs proximal to the eighth codon of the gene. All the restriction sites have been phased to three different reading frames.

Two series of vectors have been constructed. The first series (YEp) has two origins of replication (*ori*), i.e., of the yeast 2μ circle and of the ColE1 plasmid of *E. coli*, and can therefore replicate autonomously in both organisms. These shuttle vectors also have the Ap^R gene of *E. coli* and either the yeast *LEU2* or *URA3* genes to allow for selection of both *E. coli* and yeast transformants. The second series of vectors (YIp) are identical in all respects to the YEp vectors except that they lack the 2μ *ori*. The YIp vectors can be used to integrate *lacZ* fusions into yeast chromosomal DNA. None of the vectors express β-galactosidase (βGal) in yeast or *E. coli* in the absence of inserted yeast promoter sequences. The 5′-nontranslated sequences and parts of the coding sequences of various yeast genes have been cloned into representative *lacZ* fusion vectors. In-frame gene fusions can be detected by βGal activity when either yeast or *E. coli* clones are plated on media containing XGal indicator. Quantitative determinations of promoter activity were made by colorimetric assay of βGal activity in whole cells. Fusion of the yeast *CYC1* gene to *lacZ* in one of the vectors allowed detection of regulated expression of this gene when cells were grown under conditions of catabolite repression or derepression.

* To whom correspondence and requests for plasmids, sequences and reprints should be addressed.

Abbreviations: Ap, ampicillin; βGal, β-galactosidase; bp, base pair(s); kb, 1000 bp; MCR, multiple cloning region; LB, M63, WO, YPD, see MATERIALS AND METHODS, section a; nt, nucleotide(s); ONPG, *o*-nitrophenyl-β-*d*-galactoside; *ori*, origin of DNA replication; PA, polyacrylamide; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; XGal, 5-bromo-4-chloro-indolyl-β-D-galactoside; 2μ, yeast 2μ circular plasmid DNA.

INTRODUCTION

Fusion of DNA sequences to the *lacZ* gene of *E. coli* provides a convenient means of studying prokaryotic and eukaryotic promoters and regulatory elements (Bassford et al., 1978; Guarente, 1983; Rose and Botstein, 1983). The ability of *Saccharomyces cerevisiae* to synthesize active βGal has been extensively exploited to identify and delineate regulatory sequences in the 5′ non-coding regions of yeast genes (for example see Rose et al., 1981; Guarente and Ptashne, 1981; Guarente and Mason, 1983; Guarente et al., 1984; Struhl, 1982; Lucchini et al., 1984). The βGal fusions employed in such studies involve ligation of the 5′ upstream regions and part of the coding region of a yeast gene to the *lacZ* gene of *E. coli* lacking the promoter sequences, the translational signals and the first seven codons.

To simplify the construction of *lacZ* fusions in yeast we have developed a set of vectors capable of accepting DNA fragments compatible with all restriction enzyme recognition sites present in the multiple cloning region of the plasmid pUC18 (Yanisch-Perron et al., 1985). These yeast/*E. coli* shuttle vectors contain the *lacZ* gene starting from the eighth codon fused to the multiple cloning region of pUC18 with either the *Hin*dIII or the *Eco*RI site proximal to the *E. coli* gene. The restriction sites of the multiple cloning region occur in all three reading frames with respect to the *lacZ* coding sequence. Two types of vectors have been constructed. The first type, designated by the prefix YEp, contains sequences allowing autonomous replication in *E. coli* and in yeast. These vectors also contain the *E. coli* β-lactamase gene to confer Ap resistance, and either the yeast *URA3* or *LEU2* gene to permit prototrophic selection of transformants. The second set of vectors, designated by the prefix YIp, are identical to the YEp vectors except that the yeast 2μ circle sequence necessary for autonomous replication in yeast has been deleted. These vectors can be used to integrate *lacZ* fusions into yeast chromosomal DNA. The vectors have been shown to express βGal in yeast in the presence but not absence of DNA inserts with appropriate transcriptional and translational signals.

MATERIALS AND METHODS

(a) Media, strains and transformations

Non-selective medium for yeast (YPD) contained 1% yeast extract, 2% peptone and 2% glucose. Selective medium for yeast (WO) contained 0.67% yeast nitrogen base minus amino acids and 2% glucose supplemented as required with tryptophan, uracil, histidine, adenine and leucine at 25 μg/ml. *E. coli* was grown in LB medium (Davis et al., 1980) supplemented with 40 μg Ap/ml when required for selection of plasmids. E medium (Davis et al., 1980) supplemented as required was used for selection of specific markers in *E. coli*. Solid media contained 2% or 1.5% agar for growth of yeast and *E. coli*, respectively. *S. cerevisiae* strain W303-1B (α *leu2*-3,112 *his3*-11,15 *ade2*-1 *ura3*-1 *trp1*-1 *can1*-100 [cir$^+$]) obtained from R. Rothstein, College of Physicians and Surgeons, Columbia University, New York, NY, was transformed with autonomously replicating plasmids by the method of Beggs (1978). Transformants were selected on minimal glucose media lacking either uracil or leucine but supplemented for the other auxotrophic requirements of W303-1B. *E. coli* strain RR1 (*proA*, *leuB*6, *lacY*, *galK*2, *xyl*-5, *mtl*-1, *ara*-14, *rpsL*20, *supE*44, *hsdS*, λ$^-$) was used for maintenance of plasmids and for selection of plasmids containing the yeast *LEU2* gene. *E. coli* strain MC1066 (F$^-$, *Δlac*X74, *hsdR*, *rpsL*, *galU*, *galK*, *trpC*9830, *leuB*6, *pyrF*::Tn*5*) was used for selection of plasmids containing the yeast *URA3* gene. The βGal-deficient *E. coli* strain MC1009 (*araD*139, *Δlac*X74, *Δara-leu*7967, *galU*, *galK*, *strA*, *recA*56, sr1::Tn*10*, *relA*, *spoT*) was used to test expression of the plasmid copy of *lacZ* in *E. coli* by plating on LB medium supplemented with 50 μg XGal/ml. All bacterial transformations were by the $CaCl_2$ procedure (Cohen et al., 1972). βGal activity was tested in yeast by plating on M63 salts medium supplemented with 40 μg XGal/ml (Guarente, 1983).

(b) Miscellaneous procedures

Standard techniques were used for preparation of recombinant plasmids from *E. coli*, restriction enzyme digestions, agarose gel electrophoresis, isolation of restriction fragments from agarose gels,

ligation of restriction fragments and screening of transforming DNAs (Maniatis et al., 1982). Fragments with protruding 5′ ends were converted to blunt-ended fragments using the PolIk (Maniatis et al., 1982). Controlled exonucleolytic digestion of double-stranded DNA was accomplished by treatment of approx. 1 μg DNA with 50 units of S1 nuclease in 50 μl of 30 mM NaCl, 1 mM $ZnCl_2$, 35 mM sodium acetate, pH 4.75, for 5 min at 37 °C. The reaction was stopped by adding 100 μl of 100 mM Tris · HCl, pH 7.5, 10 mM EDTA, followed by phenol extraction. This S1 nuclease treatment resulted in the loss of 7–14 nt from each end of the molecules. DNA sequences were determined by the method of Maxam and Gilbert (1977). Quantitative determination of βGal activity in yeast cells was performed by measuring hydrolysis of ONPG as described (Guarente, 1983).

RESULTS AND DISCUSSION

(a) Construction of *lacZ* fusion vectors for expression of βGal in yeast: YEp353, YEp354 and YEp355

The βGal fusion vector pMC1403 (Casadaban et al., 1980) was modified by Minton (1984) to allow fusion to the *lacZ* structural gene in three reading frames. Three plasmids designated pNM480, pNM481, and pNM482, all contain the multiple cloning region of the plasmid pUC8 (Vieira and Messing, 1982) upstream from the *lacZ* gene, with a phase correction between the *Hin*dIII site and the eighth codon of *lacZ* (Minton, 1984). The availability of these plasmids suggested a simple means for introducing the three different pUC8/*lacZ* sequences into a yeast shuttle vector. For this purpose we chose the episomal plasmids YEp351 and YEp352 (Hill et al., 1986) both of which contain the entire pUC18 sequence, the yeast 2μ origin of replication, and the wild-type *LEU2* or *URA3* genes, respectively. Initially the 3.15-kb *Eco*RI-*Dra*I fragment of each pNM vector was ligated separately to YEp351 or YEp352 from which 215 bp between the *Eco*RI and *Nar*I sites had been removed (Fig. 1). The resultant plasmids YEp353A, YEp354A, YEp355A and YEp363A were capable of replicating in *E. coli* and in yeast and of complementing the *leu2* or *ura3* mutations of an appropriately marked yeast strain. Some of the plasmids, however, expressed βGal activity when yeast or *E. coli* transformants were plated in the presence of XGal. The synthesis of βGal is probably due to the presence of the *lac* promoter and an ATG start codon upstream from the multiple cloning regions of YEp351 and YEp352.

The *lac* promoter region was removed from YEp352 as shown in Fig. 1. YEp352 was digested to completion with *Pvu*II to eliminate the entire *lacZ′* region of pUC18 as well as the operator/promoter and part of the *lacI* gene. The digestion mixture was then briefly treated with S1 nuclease to remove an ATG codon located immediately 5′ of the *Pvu*II site within the *lacI* gene. The 8.0-kb vector band was purified and was ligated to a blunt-ended *Eco*RI linker; the sequence of this linker was 5′–CCCGGATTCGGG–3′. Several different plasmids containing the *Eco*RI site were partially sequenced to determine the effects of digestion with S1 nuclease. The plasmid YEp352E was ascertained to have lost 6 nt, including the ATG sequence on the 5′ side of the *lacI* coding sequence.

The pUC8/*lacZ* sequences from the previous set of vectors YEp353A, YEp354A and YEp355A were transferred into YEp352E (Fig. 1). The 3.9-kb *Eco*RI-*Nco*I fragments of YEp353A, YEp354A and YEp355A containing the multiple cloning region, the *lacZ* gene and part of *URA3* was purified from each vector and ligated to the large *Eco*RI-*Nco*I fragment of YEp352E. Following transformation of *E. coli* RR1 with the ligation mixtures, Ap-resistant clones were screened for plasmids having the pUC8/*lacZ* sequences and the reconstituted *URA3* gene. These plasmids designated YEp353, YEp354 and YEp355 did not express βGal activity when transformed into either yeast or *E. coli* (see section e below). The disposition of the restriction sites in the pUC8 multiple cloning region with respect to the *lacZ* reading frame was verified by nt sequence analysis (Table I). The complete nt sequences of the YEp vectors containing *URA3* were compiled from the known sequences of YEp352 and the pNM vectors (Table II). All restriction sites of the multiple cloning region occur once in these constructs, with the exception of *Sac*I which is also present in the coding sequence of *lacZ*.

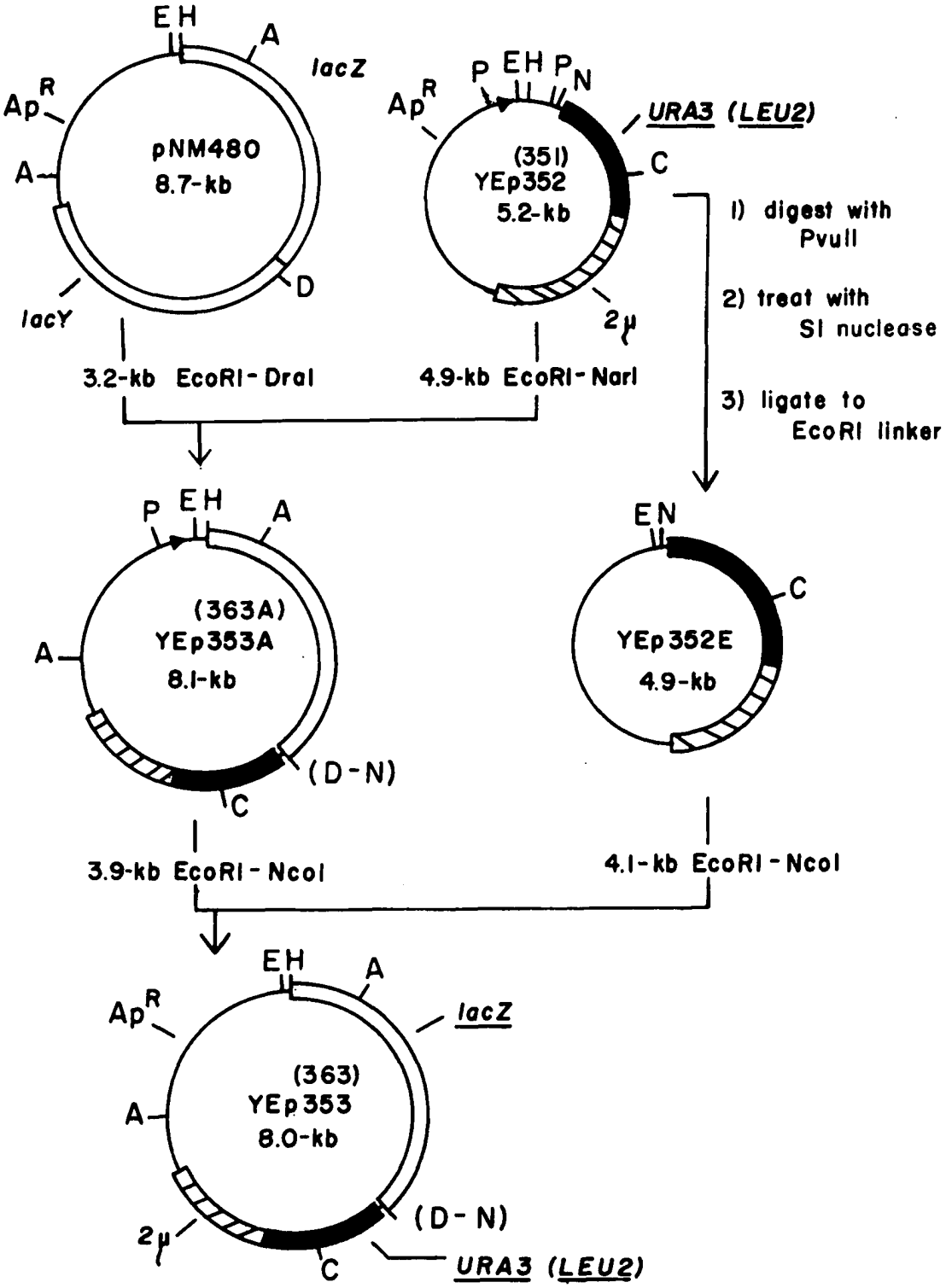


Fig. 1. Construction of YEp353 and YEp363. Single line, pUC8 or pUC18 sequence. Open box, *lacZ* or *lacY* sequence, solid box, *URA3* or *LEU2* sequence, cross-hatched box: 2μ circle sequence. The figure is drawn to scale for the *URA3* containing vectors except for the 0.3-kb region spanning the *Pvu*II sites which has been expanded for detail. The indicated vector sizes apply to scaled vectors only. The arrangement of genes in the corresponding *LEU2* containing vectors YEp351, YEp363A and YEp363 (indicated by numbers in parentheses) is also represented by these figures although in this case scale is no longer maintained. Restriction sites are indicated for *Aat*II (A), *Dra*I (D), *Eco*RI (E), *Hin*dIII (H), *Nar*I (N), *Nco*I (C), and *Pvu*II (P). The *Nco*I site is present only in the vectors containing *URA3*. (D-N) indicates the ligated junction of free ends created by cleavage with *Nar*I and *Dra*I, where neither restriction site was recreated. Arrowheads indicate the *lac* promoter. YEp363 was formed by ligating the 2.7-kb *Aat*II fragment of YEp353 to the 5.2-kb *Aat*II fragment of YEp363A. YEp vectors with MCRs in the other two reading frames were constructed by repeating the manipulations diagrammed here using pNM481 and pNM482 as the source of the *lacZ* gene.

TABLE I

Structure of the multiple cloning regions (MCR) of the *lacZ* fusion vectors

Vector	Sequence and reading frame of multiple cloning region[a]
YEp353, YEp363	5'-GA ATT CCC GGG GAT CCG TCG ACC TGC AGC CAA GCT TGC GAT CCC-3'
YEp354, YEp364	5'-GAA TTC CCG GGG ATC CGT CGA CCT GCA GCC AAG CTT GCT CCC-3'
YEp355, YEp365	5'-G AAT TCC CGG GGA TCC GTC GAC CTG CAG CCA AGC TTC GAT CCC-3'
YEp356, YEp366	5'-GA ATT CGA GCT CGG TAC CCG GGG ATC CTC TAG AGT CGA CCT GCA GGC ATG CAA GCT TGC GAT CCC-3'
YEp357, YEp367	5'-GAA TTC GAG CTC GGT ACC CGG GGA TCC TCT AGA GTC GAC CTG CAG GCA TGC AAG CTT GCT CCC-3'
YEp358, YEp368	5'-G AAT TCG AGC TCG GTA CCC GGG GAT CCT CTA GAG TCG ACC TGC AGG CAT GCA AGC TTC GAT CCC-3'
YEp356R, YEp366R	5'-AAG CTT GCA TGC CTG CAG GTC GAC TCT AGA GGA TCC CCG GGT ACC GAG CTC GAA TTC CCA GCT TGC GAT CCC-3'
YEp357R, YEp367R	5'-A AGC TTG CAT GCC TGC AGG TCG ACT CTA GAG GAT CCC CGG?GTA CCG AGC TCG AAT TCC CAG CTT GCT CCC-3'
YEp358R, YEp368R	5'-AA GCT TGC ATG CCT GCA GGT CGA CTC TAG AGG ATC CCC GGG TAC CGA GCT CGA ATT CCC AGC TTC GAT CCC-3'

Cleavage position[b]

Restriction site	YEp363 YEp353	YEp364 YEp354	YEp365 YEp355	YEp366 YEp356	YEp367 YEp357	YEp368 YEp358
5'-*Eco*RI	2	1	3	2*	1	3
*Sac*I	—	—	—	3*	2	1
*Kpn*I	—	—	—	3*	2	1
*Sma*I	3	2	1	2*	1	3
*Bam*HI	3	2	1	2*	1	3
*Xba*I	—	—	—	2*	1	3
*Sal*I	3	2	1	2	1	3
*Pst*I	1	3	2	3	2	1
*Sph*I	—	—	—	3	2	1
3'-*Hin*dIII	2	1	3	2	1	3

Cleavage position[b]

Restriction site	YEp366R YEp356R	YEp367R YEp357R	YEp368R YEp358R
5'-*Hin*dIII	1	3	2*
*Sph*I	2	1	3*
*Pst*I	2	1	3*
*Sal*I	1	3	2*
*Xba*I	1	3	2*
*Bam*HI	1	3	2
*Sma*I	2	1	3
*Kpn*I	2	1	3
*Sac*I	2	1	3
3'-*Eco*RI	1	3	2

[a] The sequences begin at the first nt of the MCR and end with the CCC codon specifying the eighth amino acid of βGal. The sequences of all vectors with a given selectable marker are identical outside of the region shown.

[b] Sites where a given enzyme will cleave within a triplet, with '3' meaning between triplets. All restriction sites listed except for *Sac*I are unique in YEp353-YEp358R. All restriction sites listed except for *Eco*RI, *Kpn*I and *Sac*I are also unique in YEp363-YEp368R. The MCR of each integrative YIp vector is identical to the MCR of the episomal YEp vector with the corresponding numerical suffix. Cloning into any site marked with an asterisk will lead to a TAG stop codon within the downstream *Xba*I site of the MCR.

TABLE II

Restriction enzyme recognition sites in YEp356[a]

Enzyme	No. of sites	Position of sites													
*Aat*II	2	672	5979												
*Acc*I	3	34	3278	3783											
*Acy*I	5	672	2200	2372	5979	6361									
*Afl*III	6	1355	2135	2560	2812	3928	7790								
*Aha*III	3	6322	7014	7033											
*Alu*I	29	8	53	144	255	435	1990	2290	2701	3064	3258	4051	4101	4330	4639
		4764	5048	5450	5545	5620	5634	5850	5869	6548	6611	6711	7232	7489	7625
		7851													
*Apa*I	1	3733													
*Asu*I	14	164	1594	2480	2742	2880	3733	3734	3955	4952	5922	6538	6760	6777	6856
*Asu*II	1	3843													
*Ava*I	4	17	1391	2834	4855										
*Ava*II	5	1594	3955	4952	6538	6760									
*Ava*III	2	3284	4849												
*Bam*HI	1	22													
*Bcl*I	1	1399													
*Bgl*I	3	199	2320	6778											
*Bin*I	9	22	1462	1754	2795	6255	6576	7040	7138	7224					
*Bse*PI	1	1551													
*Bst*XI	2	2266	2883												
*Cau*II	15	17	18	1487	1577	1936	2113	2173	3120	3505	4512	5829	5864	6365	6716
		7412													
*Cfr*I	5	524	1522	3319	6509	7951									
*Cla*I	1	876													
*Dde*I	16	277	558	2217	3066	3378	3615	4211	4636	4761	4885	5740	5975	6401	6941
		7107	7516												
*Eco*RI	1	1													
*Eco*RII	14	96	223	527	1620	2188	2455	2489	3034	3730	4368	4462	7630	7643	7764
*Eco*RV	2	1164	3921												
*Gdi*II	5	524	1522	3319	6509	7951									
*Hae*I	8	449	1427	2687	3126	3673	7314	7766	7777						
*Hae*II	20	214	508	945	1434	1886	2009	2109	2184	3069	4427	4611	4674	4736	4799
		4861	5500	5519	5597	7546	7916								
*Hae*III	25	167	281	450	525	1257	1428	1523	2481	2492	2688	2743	2880	3127	3320
		3674	3734	5923	6510	6777	6857	7315	7749	7767	7778	7952			
*Hga*I	15	455	904	1446	1628	2201	2225	2238	2372	2861	3478	5490	5803	6361	7111
		7689													
*Hgi*AI	9	7	1989	2398	2513	4219	5733	6230	6315	7476					
*Hgi*CI	6	13	233	245	849	1301	6949								
*Hgi*EII	3	236	3223	7203											
*Hgi*JII	5	7	1989	3038	3640	3733									
*Hin*cII	7	34	477	1101	2929	3467	3603	4561							
*Hin*dIII	1	52													
*Hin*fI	20	32	392	959	1091	1310	1422	3030	4259	4632	4694	4757	4819	5014	5147
		5240	6903	7420	7816	7891	7956								
*Hin*fIII	12	3	209	459	958	1092	1365	1649	1961	2867	3753	4258	5433		
*Hpa*I	3	477	1101	4561											
*Hpa*II	30	18	231	249	577	1341	1488	1578	1936	2103	2113	2173	2597	2609	2676
		3073	3120	3505	4513	5463	5830	5864	6365	6607	6717	6784	6818	7222	7412
		7438	7585												
*Hph*I	15	596	828	867	1388	1809	2037	2392	3367	5890	5899	6183	6198	6424	6820
		7047													
*Kpn*I	1	13													

(TABLE II, continued)

Enzyme	No. of sites	Position of sites													
*Mae*I	11	29	3679	4069	4425	5041	5292	5543	5547	6709	7044	7297			
*Mae*III	30	82	102	322	348	406	520	616	670	793	807	906	1108	1359	1669
		1953	2223	3360	3528	3691	4337	5455	5854	6242	6430	6583	6641	6972	7255
		7371	7434												
*Mbo*I	30	23	60	174	270	636	900	1326	1380	1400	1463	1476	1640	1755	1833
		2385	2796	3430	6220	6556	6273	6531	6577	6595	6936	7041	7053	7131	7139
		7150	7225												
*Mbo*II	30	159	272	633	1563	2830	2976	3327	3408	3540	3593	3697	3841	3846	4119
		4178	4235	4438	4691	4967	4995	5150	5180	5493	5692	6108	6217	6295	7050
		7141	7912												
*Mlu*I	3	1355	2135	2560											
*Mnl*I	27	26	162	279	301	590	762	1041	1096	1122	2046	2216	3371	3406	3676
		3899	5316	5883	5925	6536	6742	6872	6953	7353	7620	7677	7903	7936	
*Mst*I	4	193	5110	5623	6677										
*Mst*II	1	276													
*Nco*I	1	3902													
*Nde*I	2	3009	4190												
*Nla*III	27	47	717	1128	1139	1298	2161	2813	2986	3130	3283	3549	3649	3903	3929
		4038	4110	4172	5687	5874	5958	6063	6456	6492	6570	6580	7071	7791	
*Nsp*BII	14	143	845	1227	1635	2025	2376	2403	2700	2789	3063	5798	6264	7205	7450
*Nsp*CI	6	46	2812	3282	3928	5873	7790								
*Pss*I	2	3954	5921												
*Pst*I	1	40													
*Pvu*I	5	173	899	1379	1832	6530									
*Pvu*II	3	143	2700	3063											
*Rsa*I	13	14	756	1235	1547	2133	2825	3167	3610	3798	3861	3994	5744	6420	
*Sac*I	2	7	1989												
*Sal*I	1	34													
*Sca*I	2	3797	6419												
*Sdu*I	14	7	1989	2175	2398	2513	3038	3640	3733	3895	4219	5733	6230	6315	7476
*Sma*I	1	17													
*Sna*I	3	2816	3278	3783											
*Sna*BI	1	5203													
*Sph*I	1	46													
*Ssp*I	3	1281	5439	6095											
*Stu*I	1	3673													
*Taq*I	16	5	35	877	925	1094	1474	1948	2287	2461	2998	3844	4106	5238	5320
		6248	7692												
*Tth*III	7	4023	4624	4749	5417	7177	7183	7216							
*Xba*I	1	28													
*Xho*II	9	22	2795	3429	6255	6272	7040	7052	7138	7149					
*Xmn*I	2	5172	6298												

[a] The vector contains 7966 bp numbered on the coding strand of *lacZ*, with nt 1 defined as the first nt of the MCR. Differences between YEp356 and other YEp vectors containing *URA3* are shown in Table I. The MCR includes nt 1–62, the *lacZ* sequence includes nt 63–3175, the *URA3* sequence includes nt 3231–4333, the 2μ sequence includes nt 4334–5728, and the pUC sequence includes nt 5729–7966 and 3176–3230.

(b) Construction of YEp356, YEp357, YEp358, YEp356R, YEp357R and YEp358R

Additional unique restriction sites were introduced into the *lacZ* fusion vectors upstream from the βGal gene by replacement of the multiple cloning region of pUC8 with that of pUC18. The multiple cloning region of pUC18 was isolated as a 56-bp *Eco*RI-*Hin*dIII fragment and ligated separately to YEp353, YEp354 and YEp355 digested with *Eco*RI and *Hin*dIII. The resultant recombinant plasmids YEp356, YEp357, and YEp358, contain unique *Sph*I, *Kpn*I, and *Xba*I recognition sites upstream from the *lacZ* gene in addition to those present in the pUC8 multiple cloning region (Table I). The three vectors also contain a *Sac*I recognition site available for cloning in frame gene fusions, although this site is also present once in the *lacZ* sequence. The phasing of the reading frames of the unique cloning sites of YEp356-YEp358 relative to *lacZ* were verified by nt sequence analysis (Table I).

To facilitate cloning of yeast genomic fragments defined by different upstream restriction sites, the multiple cloning regions of YEp356, YEp357 and YEp358 were inverted with respect to *lacZ*. A blunt ended form of the pUC18 multiple cloning region was constructed from the 56 bp *Eco*RI-*Hin*dIII fragment used earlier. The fragment was first methylated using *Hpa*II methylase to protect the internal 5′–CCCGGG–3′ sequence from digestion with *Sma*I. The methylated fragment was ligated to two adaptor sequences, the *Eco*RI-*Sma*I adaptor 5′–GAATCCCGGG–3′ and the *Hin*dIII-*Sma*I adaptor 5′–AAGCTTCCCGGGA–3′. These two adaptor sequences recreate the *Eco*RI and *Hin*dIII sites. The high M_r ligation products were digested with *Sma*I and the unit-length, blunt-ended fragment with the multiple cloning region was purified on a 6% PA gel. This fragment was ligated separately to YEp356, YEp357, and YEp358 which had been digested with *Eco*RI + *Hin*dIII and been made blunt-ended by treatment with PolIk. The ligation mixture was used to transform *E. coli* strain RR1, and individual clones were screened by restriction mapping for plasmids in which the orientation of the multiple cloning region was opposite that of the parent vectors. Three plasmids designated YEp356R, YEp357R, and YEp358R were confirmed by DNA sequence analysis to have the *Eco*RI site proximal to the *lacZ* gene with the unique cloning sites in each of the three reading frames (Table I).

(c) Construction of *lacZ* fusion vectors containing *LEU2* as a selectable marker

A second set of *lacZ* fusion vectors containing the yeast *LEU2* gene as a selectable marker was constructed by transferring segments of each *URA3* containing vector to YEp363A (Fig. 1). YEp353 through YEp358R were used to prepare a 2.7-kb *Aat*II fragment containing most of the pUC18 sequence, the multiple cloning region, and the 5′ region of *lacZ*. These fragments were ligated separately to the 5.8-kb *Aat*II fragment of YEp363A containing the 3′ region of *lacZ*, the yeast *LEU2* gene, and the remainder of pUC18 (Fig. 1). The ligation mixture was used to transform *E. coli* RR1 and Ap-resistant colonies were scored for leucine prototrophy by complementation of the *leuB* mutation of *E. coli*. Plasmid DNA extracted from the Leu^+ clones was analyzed by restriction mapping to confirm reconstitution of the *lacZ* gene. The resultant plasmids YEp363-YEp368R contain MCRs with the same disposition of reading frames as the corresponding *URA3* vectors YEp353-YEp358R (Table I). The complete nt sequences of these vectors compiled from the known sequences of YEp351 and the pNM vectors shows the *Eco*RI, *Sac*I, and *Kpn*I sites are present twice while the remainder of the sites in the MCR are unique (Table III).

(d) Construction of integrative *lacZ* fusion vectors

Each of the episomal *lacZ* fusion vectors described above were converted to integrative vectors by removal of yeast 2μ circle sequences required for autonomous replication in yeast (Fig. 2). In the case of the vectors with the *URA3* gene, the 5.9-kb region from the *Aat*II site of pUC18 to the *Nco*I site in *URA3* was ligated to the 1.2-kb *Aat*II-*Nco*I fragment of the integrative vector YIp352 (Hill et al., 1986). This fragment of YIp352 supplies the sequences necessary for reconstitution of pUC18 and *URA3*, but does not contain the 2μ sequence essential for autonomous replication in yeast. A similar approach was used to construct integrative forms of the *LEU2* vectors. The region of YEp363-YEp368R from the *Kpn*I site of the *LEU2*

TABLE III

Restriction enzyme recognition sites in YEp366[a]

Enzyme	No. of sites	Position of sites													
*Aat*II	2	672	6437												
*Acc*I	2	34	3602												
*Acy*I	5	672	2200	2372	6437	6819									
*Afl*II	1	4479													
*Afl*III	6	1355	2135	2560	2812	4057	8248								
*Aha*III	3	6780	7472	7491											
*Alu*I	25	8	53	144	255	435	1990	2290	2701	3064	3553	4220	4712	5279	5404
		5688	6090	6308	6327	7006	7069	7169	7690	7947	8083	8309			
*Asu*I	15	164	1594	2480	2742	2880	4232	4264	4560	4904	5592	6380	6996	7218	7265
		7314													
*Asu*II	2	4673	5161												
*Ava*I	4	17	1391	2834	5495										
*Ava*II	7	1594	4232	4560	4904	5592	6996	7218							
*Ava*III	1	5489													
*Bam*HI	1	22													
*Bcl*I	1	1399													
*Bgl*I	3	199	22320	7236											
*Bin*I	10	22	1462	1754	2795	4108	6713	7034	7498	7596	7682				
*Bse*PI	1	1551													
*Bst*EII	1	4759													
*Bst*XI	3	2266	2883	4080											
*Cau*II	14	17	18	1487	1577	1936	2113	2173	3120	3921	6287	6322	6823	7174	7870
*Cfr*I	5	524	1522	4276	6967	8409									
*Cla*I	2	876	4643												
*Dde*I	16	277	558	2217	3066	44784	5153	5186	5276	5401	5525	6198	6433	6859	7399
		7565	7974												
*Eco*RI	2	1	4157												
*Eco*RII	13	96	223	527	1620	2188	2455	2489	3034	4017	4764	8088	8101	8222	
*Eco*RV	2	1164	4046												
*Gdi*II	5	524	1522	4276	6967	8409									
*Hae*I	10	449	1427	2687	3126	3826	3982	4207	7772	8224	8235				
*Hae*II	19	214	508	945	1434	1886	2009	2109	2184	3069	5106	5251	5314	5376	5439
		5501	6140	6159	8004	8374									
*Hae*III	29	164	281	450	525	1257	1428	1523	2481	2492	2688	2743	2880	3127	3249
		3827	3983	4208	4265	4277	4993	6381	6968	7235	7315	7773	8207	8225	8236
		8410													
*Hga*I	8	7	1989	2398	2513	6191	6688	6773	7934						
*Hgi*CI	8	13	233	245	849	1301	3457	4547	7407						
*Hgi*EII	4	236	3734	3953	7661										
*Hgi*JII	3	7	1989	3038											
*Hin*cII	7	34	477	1101	2929	3899	4921	5201							
*Hin*dIII	1	52													
*Hin*fI	26	32	392	959	1091	1310	1422	3030	3413	3696	4122	4289	4455	5093	5159
		5272	5334	5397	5459	5654	5787	5880	7361	7878	8274	8349	8414		
*Hin*fIII	14	3	209	459	958	1092	1365	1649	1961	2867	3414	4159	5086	5160	6073
*Hpa*I	3	477	1101	5201											
*Hpa*II	33	18	231	249	577	1341	1488	1578	1936	2103	2113	2173	2597	2609	2676
		3073	3120	3234	3921	4545	4986	4958	4991	6103	6288	6322	6823	7065	7175
		7276	7680	7870	7896	8043									
*Hph*I	22	596	828	867	1388	1809	2037	2392	3286	3741	3777	4050	4067	4341	4761
		4999	6348	6357	6641	6656	6882	7278	7505						
*Kpn*I	2	13	4547												
*Mae*I	8	29	4105	5681	5932	6183	7167	7502	7755						

(TABLE III, continued)

Enzyme	No. of sites	Position of sites													
*Mae*III	28	82	102	322	348	406	520	616	670	793	807	906	1108	1359	1669
		1953	2223	4469	4760	6095	6312	6700	6888	7041	7099	7430	7713	7829	7892
*Mbo*I	32	23	60	174	270	636	900	1326	1380	1400	1463	1476	1640	1755	1833
		2385	2796	3231	4109	4778	6678	6714	6731	6989	7035	7053	7394	7499	7511
		7589	7597	7608	7683										
*Mlu*I	3	1355	2135	2560											
*Mst*I	3	193	5750	7135											
*Mst*II	1	276													
*Nde*I	1	3009													
*Nla*III	26	47	717	1128	1139	1298	2161	2813	2986	3130	3425	3939	4058	4112	4255
		4268	4473	4955	6332	6416	6521	6914	6950	7028	7038	7529	8249		
*Nsp*BII	14	143	845	1227	1635	2025	2376	2430	2700	2789	3063	6256	6722	7663	7908
*Nsp*CI	5	46	2812	4057	6331	8248									
*Pss*I	2	4903	6379												
*Pst*I	1	40													
*Pvu*I	5	173	899	1379	1832	6988									
*Pvu*II	3	143	2700	3063											
*Rsa*I	16	14	756	1235	1547	2133	2825	3167	3237	3675	3756	3947	4410	4490	4548
		6202	6878												
*Sac*I	2	7	1989												
*Sal*I	1	34													
*Sca*I	1	6877													
*Sdu*I	11	7	1989	2175	2398	2513	3038	3458	6191	6688	6773	7934			
*Sma*I	1	17													
*Sna*I	1	2816													
*Sna*BI	1	5843													
*Sph*I	1	46													
*Ssp*I	5	1281	3475	4915	6079	6553									
*Taq*I	19	5	35	877	925	1094	1474	1948	2287	2461	2998	4644	4674	4680	5085
		5162	5878	5960	6706	8150									
*Tth*III	9	4054	4792	4943	5264	5389	6057	7635	7641	7674					
*Xba*I	1	28													
*Xho*II	9	22	2795	3230	6713	6730	7498	7510	7596	7607					
*Xmn*I	4	3613	4157	5812	6756										

[a] The vector contains 8424 bp numbered as described in Table II. Differences between YEp366 and other YEp vectors containing *LEU2* are shown in Table I. The MCR includes nt 1–62, the *lacZ* sequence includes nt 63–3175, the *LEU2* sequence includes nt 3231–5202, the 2μ sequence includes nt 5203–6186, and the pUC sequence includes nt 6187–8424 and 3176–3230.

gene to the *Sca*I site of the *lacZ* gene was replaced with the corresponding *Kpn*I-*Sca*I region of the integrative vector YIp351 (Hill et al., 1986) to yield a set of vectors with a deletion in the yeast 2μ sequence. The integrative vectors containing *URA3* as a selectable marker are designated YIp353-YIp358R and those containing *LEU2* are designated YIp363-YIp368R (Table I).

(e) Properties of the *lacZ* fusion vectors

Each of the autonomously replicating plasmids constructed in this study were used to transform yeast strain W303-1B to leucine or uracil independence. All vectors transformed yeast at the high frequency seen for other episomal plasmids containing the 2μ circle origin of replication (Broach, 1983). Greater than 80% of the segregants tested from several different transformants retained the appropriate prototrophic marker after growth for

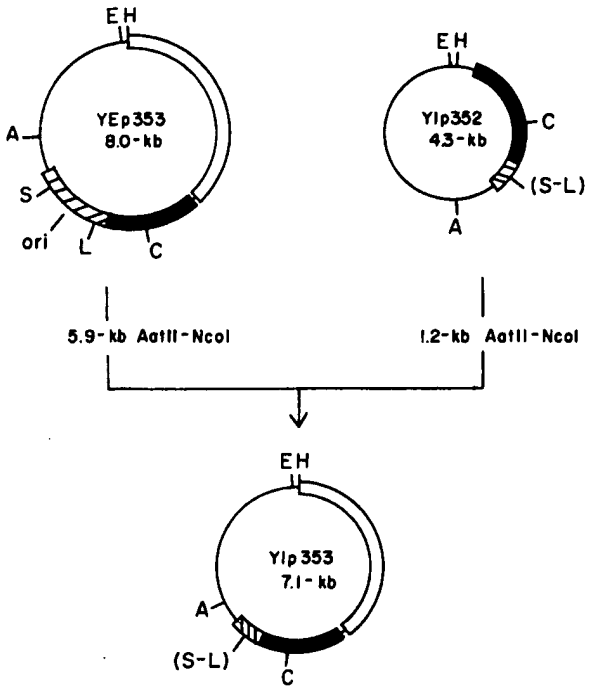


Fig. 2. Construction of the integrative vector YIp353. YIp352 was derived from YEp352 by removal of 2μ sequence necessary for autonomous replication in yeast (Hill et al., 1986). Symbols are as in Fig. 1. Additional restriction sites are indicated for *Hpa*I (L) and *Ssp*I (S). (S-L) indicates the ligation junction of free ends created by cleavage with *Ssp*I and *Hpa*I, where neither restriction site was recreated. Not all restriction sites shown in this drawing are unique in the vectors. Symbol *ori* indicates the position of the yeast 2μ origin of replication.

30–40 generations in non-selective medium, indicating that the episomal plasmids are retained at a high copy number.

The ability of the episomal plasmids to express βGal activity was tested in the yeast transformants and in the *ΔlacZ E. coli* strain MC1009. In the absence of yeast promoter sequences ligated into the multiple cloning region none of the plasmids induced the characteristic blue color indicative of βGal activity when yeast or *E. coli* transformants were grown on plates containing XGal. Various segments of yeast DNA have been cloned into the appropriate vectors to create in-frame gene fusions to *lacZ*. Among the yeast genes tested are *MRP2* coding for a mitochondrial ribosomal protein (Myers and Tzagoloff, 1986), *MSD* coding for the mitochondrial aspartyl tRNA synthetase (A. Gampel and A. Tzagoloff, unpublished results), *CPA2* coding for the large subunit of carbamyl phosphate synthetase (Lusty et al., 1983) and *CYC1* coding for apo-iso-1-cytochrome *c* (Montgomery et al., 1978). Fusion of the upstream regions and part of the coding sequence of these genes to the *lacZ* gene of different episomal plasmids described here allowed detection of βGal activity when either *E. coli* or yeast transformants were plated in the presence of XGal. Quantitative determinations of βGal activity were made by measuring hydrolysis of ONPG by yeast cells grown in liquid cultures. The βGal activity expressed from the plasmids differed depending on the particular cloned yeast promoter. The βGal activity present in a particular transformant also differed depending on growth conditions, indicating that promoter activity could be assayed in the YEp vectors by measurement of *lacZ* expression. For example, the βGal activity derived from one episomal plasmid containing the 5′-nontranslated region and first two codons of *CYC1* fused to *lacZ* was five times greater in cells grown under conditions of catabolite derepression (ethanol) than when the same transformant was grown under conditions of catabolite repression (glucose). These results were comparable to those obtained using an isogenic transformant containing pLGΔ312, a βGal fusion construct used previously for studies of regulation of *CYC1* (Guarente and Mason, 1983; Guarente et al., 1984).

The upstream region of the *CYC1* gene was also cloned into the integrative plasmid YIp356R. The recombinant plasmid was linearized within the *URA3* gene by digestion with *Nco*I and used to transform yeast strain W303-1B to uracil independence. Transformants were isolated in which greater than 99% of the segregants retained both the *URA3* gene and *lacZ* activity after growth for 30–40 generations in non-selective media indicating that the *lacZ* fusion had stably integrated into the yeast genome.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Dr. Nigel Minton for supplying the plasmids pNM480, pNM481 and pNM482. We also thank Ms. Anna Akai for providing expert technical assistance in this study. This research was supported by NIH grants

HL22174 and GM25846, and by Research Service Award GM09109 (to A.M.M.) from the National Institutes of Health, United States Public Health Service.

REFERENCES

Bassford, P.J., Beckwith, J., Berman, M., Brickman, E., Casadaban, M., Guarente, L., Saint-Girons, I., Sarthy, A. Schwartz, M., Shuman, H. and Silhavy, T.: Genetic fusions of the *lac* operon: a new approach to the study of biological processes. In Miller, J.H. and Reznikoff, W.S. (Eds.), The Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1978, pp. 245–261.

Beggs, J.D.: Transformation of yeast by a replicating hybrid plasmid. Nature 275 (1978) 104–109.

Broach, J.R.: Construction of high copy yeast vectors using 2μm circle sequences. Methods Enzymol. 101 (1983) 307–325.

Casadaban, M.J., Chou, J. and Cohen, S.N.: In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. 143 (1980) 971–980.

Cohen, S.N., Chang, A.C.Y. and Hsu, L.: Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. USA 69 (1972) 2110–2114.

Davis, R.W., Botstein, D. and Roth, J.R.: Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1980.

Guarente, L.: Yeast promoters and *lacZ* fusions designed to study expression of cloned genes in yeast. Methods Enzymol. 101 (1983) 181–191.

Guarente, L. and Mason, T.: Heme regulates transcription of the *CYC1* gene of *S. cerevisiae* via an upstream activation site. Cell 32 (1983) 1279–1286.

Guarente, L. and Ptashne, M.: Fusion of *Escherichia coli lacZ* to the cytochrome *c* gene of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 78 (1981) 2199–2203.

Guarente, L., Lalonde, B., Gifford, P. and Alani, E.: Distinctly regulated tandem upstream activation sites mediate catabolite repression of the *CYC1* gene of *S. cerevisiae*. Cell 36 (1984) 503–511.

Hill, J.E., Myers, A.M., Koerner, T.J. and Tzagoloff, A.: Yeast/*E. coli* shuttle vectors with multiple unique restriction sites. Yeast (1986) in press.

Lucchini, G., Hinnebusch, A., Chen. C. and Fink, G.R.: Positive regulatory interactions of the *HIS4* gene of *Saccharomyces cerevisiae*. Mol. Cell Biol. 4 (1984) 1326–1333.

Lusty, C.J., Widgren, E.E., Broglie, K.E. and Nyunoya, H.: Yeast carbamyl phosphate synthetase. Structure of the yeast gene and homology to *Escherichia coli* carbamyl phosphate synthetase. J. Biol. Chem. 258 (1983) 14466–14472.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Maxam, A.M. and Gilbert, W.: A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74 (1977) 560–564.

Minton, N.P.: Improved plasmid vectors for the isolation of translational *lac* gene fusions. Gene 31 (1984) 269–273.

Montgomery, D.L., Hall, B.D., Gillam, S. and Smith, M.: Identification and isolation of the yeast cytochrome *c* gene. Cell 14 (1978) 673–680.

Myers, A.M. and Tzagoloff, A.: *MRP1* and *MRP2*, two yeast genes coding for mitochondrial ribosomal proteins. J. Biol. Chem. (1986) submitted.

Rose, M. and Botstein, D.: Construction and use of gene fusions to *lacZ* (β-galactosidase) that are expressed in yeast. Methods Enzymol. 101 (1983) 167–180.

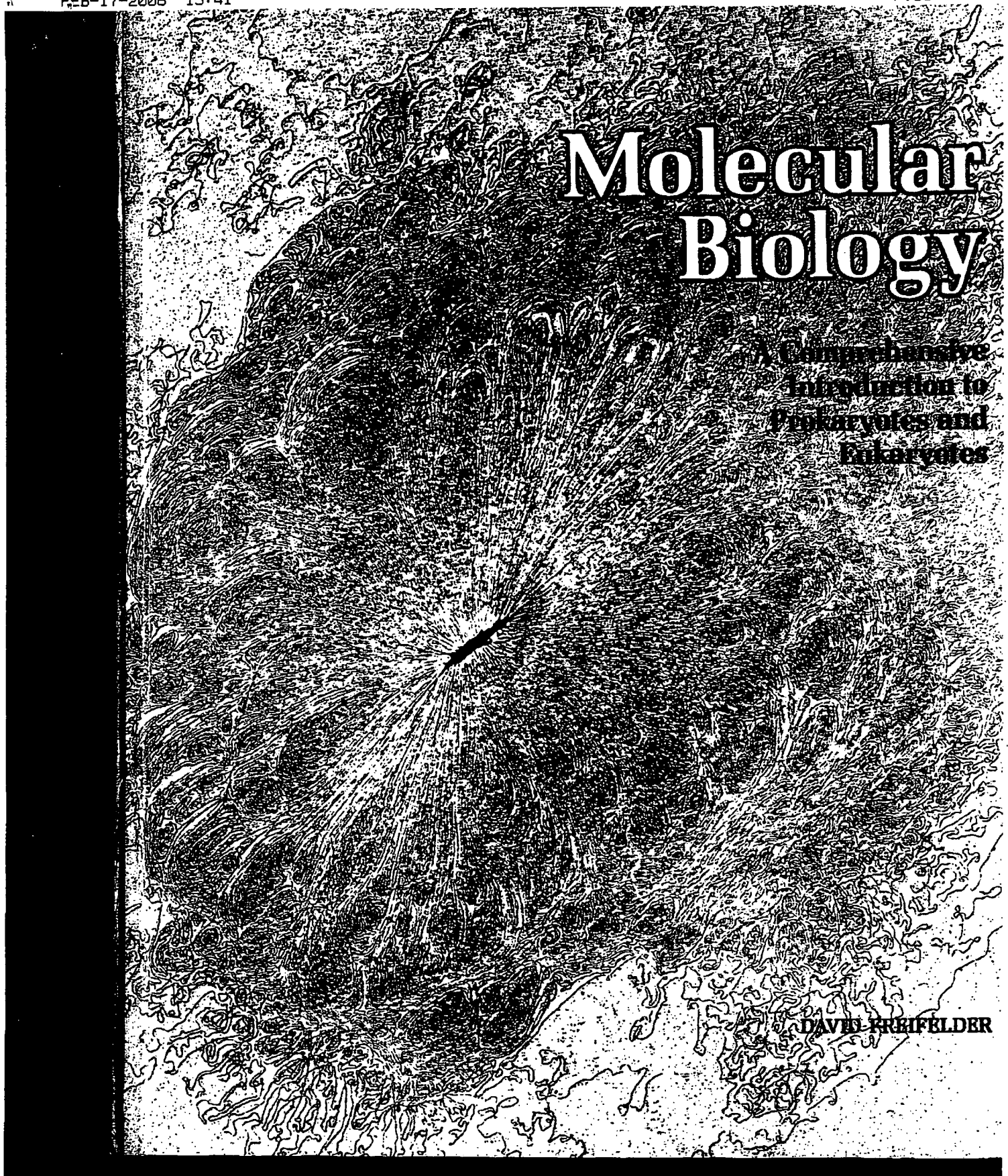
Rose, M., Casadaban, M.J. and Botstein, D.: Yeast genes fused to β-galactosidase in *Escherichia coli* can be expressed normally in yeast. Proc. Natl. Acad. Sci. USA 78 (1981) 2460–2464.

Struhl, K.: Regulatory sites for *HIS3* gene expression in yeast. Nature 300 (1982) 284–287.

Vieira, J. and Messing, J.: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19 (1982) 259–268.

Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp8 and pUC19 vectors. Gene 33 (1985) 103–119.

Communicated by G.R. Fink.



Molecular Biology

A Comprehensive Introduction to Prokaryotes and Eukaryotes

DAVID FREIFELDER

Molecular Biology

A Comprehensive Introduction to Prokaryotes and Eukaryotes

DAVID FREIFELDER
University of California, San Diego
University of Alabama
Formerly of Brandeis University

Jones and Bartlett Publishers, Inc.
BOSTON PORTOLA VALLEY

Editorial offices: Jones and Bartlett Publishers, Inc., 30 Granada Court, Portola Valley, CA 94025.

Sales and customer service offices: Jones and Bartlett Publishers, Inc., 20 Park Plaza, Boston, MA 02116.

Library of Congress Cataloging in Publication Data

Freifelder, David Michael, 1935–
Molecular biology.

Includes bibliographies and index.
1. Molecular biology I. Title.
QH506.F73 1983 574.8'8 82–17006
ISBN 0–86720–012–X

ISBN 0-86720-012-X

Publisher Arthur C. Bartlett

Book and Cover Design Hal Lockwood

Illustrator Donna Salmon, Assisted by Cyndie Clark-Huegel, Evanell Towne, John and Judy Waller, Dorothy Beebe, Kelly Solis-Navarro, and Brenda Booth

Manuscript Editor Kirk Sargent

Production Bookman Productions

Composition Typothetae

Printer and Binder Halliday Litho

Printed in the United States of America
Printing Number (last digit) 10 9 8 7 6 5 4

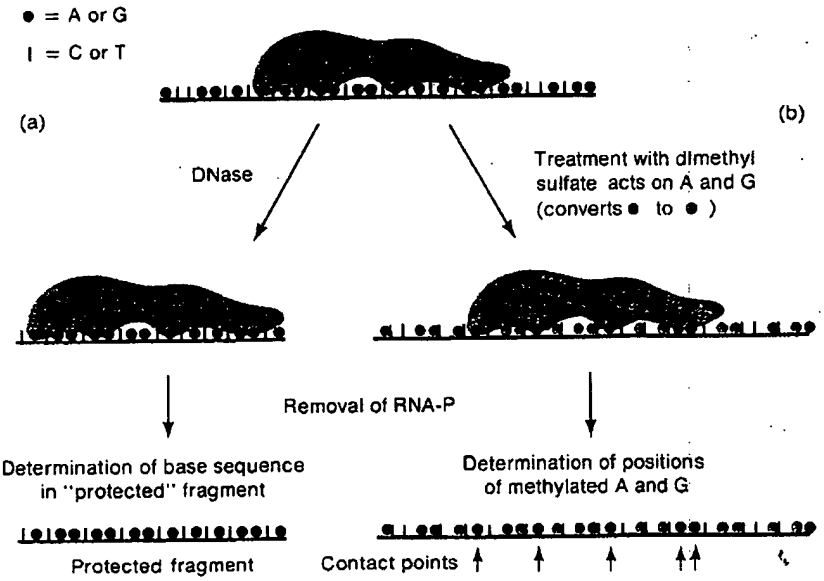


Figure 11-5
A schematic diagram of the analysis of the interaction of RNA polymerase (RNA-P) with DNA. Identification of (a) the protected region (left path of arrows) and (b) of some of the bases in contact with the enzyme (right path of arrows). For clarity, only one strand of DNA is shown, though the experiment is performed with double-stranded DNA. To locate the binding sequence the base sequence of the protected fragment is compared to that of a larger segment of DNA. Contact points are identified by comparing the methylation patterns obtained when RNA-P is either present or absent.

polymerase molecules; the number is greater when cells are growing rapidly.

Site Selection: I. The Promoter

The first step in transcription is binding RNA polymerase to a DNA molecule. Binding occurs at particular sites called **promoters**, which are specific sequences of 20–200 bases at which several interactions occur. (A promoter is also frequently defined as a region protected by RNA polymerase from digestion by endonucleases.) The existence of promoters was first demonstrated by the isolation of a particular class of Lac$^-$ mutations in *E. coli*. These mutations not only eliminate gene activity but also are noncomplementable (because they are *cis*-acting) and prevent synthesis of the RNA transcript of the *lac* gene. These mutations are called **promoter mutations.**

Several events must occur at a promoter. RNA polymerase must recognize a specific DNA sequence, attach in a proper configuration, open the DNA to gain access to the bases to be copied, and then initiate synthesis. These events are guided by the base sequence of the DNA, the polymerase σ subunit (without which the promoter is not recognized) and, for some promoters, by auxiliary proteins. The details of these events are not yet known, but the process can be broken down into three

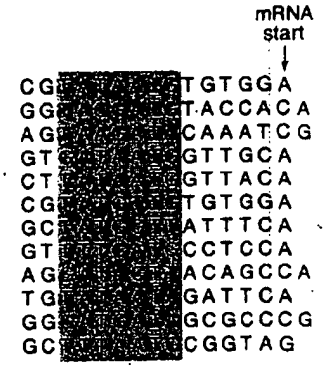


Figure 11-6
Segments of the noncoding strand of protected regions from various genes showing the common sequence of seven bases (red) known as the Pribnow box. The start point for mRNA synthesis is shown. The "conserved" T is underlined.

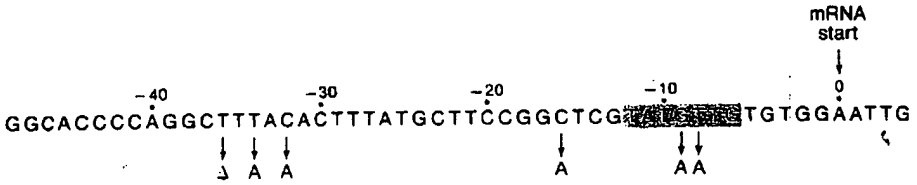


Figure 11-7
A region of the noncoding strand of the promoter for the *lac* gene showing six mutations (red arrows) that affect promoter activity; Δ means a base deletion. The Pribnow box is shaded in red. Many base changes are known; all are either in or near the Pribnow box or are clustered around base −35 and thus define an important site (see page 377).

parts—(a) template binding at a polymerase recognition site, (b) movement to an initiation site, and (c) establishment of what is termed an open-promoter complex (shown schematically later in Figure 11-9). The approach to elucidating these steps for many genes has been to isolate the DNA segment (the promoter) that is protected by RNA polymerase from DNase digestion, determine the base sequence in the segment, and look for common features in the sequences (Figure 11-6). The specific sites of contact are also determined by the dimethyl sulfate protection method. This is important because one might expect that the specific contact sites would be in the regions common to all promoters.

The RNA molecules synthesized *in vitro* from each of these promoter regions must also be sequenced if one wishes to identify the initiation sequence, which is the sequence of the first few bases that are transcribed; this sequence is just the complement of the bases at the 5′ terminus of the RNA molecule. Additional information is obtained by determining the sequence of bases in promoters having mutations that either eliminate initiation *in vivo* or change the requirements for initiation (Figure 11-7). The rationale is that if a base change affects promoter activity, that base must be contained in the promoter. This

technique has allowed researchers to identify the bases in the protected segment that are actually part of the promoter. So far, 46 promoters have been sequenced.

Site Selection: II. The Pribnow Box

Figure 11-6 shows portions of several promoter sequences in *E. coli* and *E. coli* phages (each promoter sequence is recognized by *E. coli* RNA polymerase) and their important features. In a region from five to ten bases to the left of the first base copied into mRNA is the right end of a sequence called the **Pribnow box.** All sequences found in Pribnow boxes are considered to be variants of a basic sequence TATAATG. The underscored T, at base 6 in the Pribnow box, from six to nine bases to the left of the first base transcribed (the distance depending on the distance from the Pribnow box to the transcription start point), is present in all promoters sequenced to date. It is called the "conserved T" and different sequences are usually compared by aligning conserved T's vertically, as shown in the figure. In 35 of 46 known Pribnow boxes in *E. coli*, the first two bases are TA; the variants, TG, CA, GA, and TC, retain one of the two TA bases. The Pribnow box is thought to be the sequence that orients RNA polymerase, so that synthesis proceeds from left to right (as the sequence is drawn), and the region at which the double helix opens to form the open-promoter complex (see below).

Before enough sequences were known that the conserved T was recognizable, the first base transcribed was chosen as a reference point and numbered zero. The direction of transcription was called "downstream"; all "upstream" bases, which are not transcribed, were given negative numbers starting from the zero reference. The Pribnow box is enclosed between −13 and −4, depending on the particular promoter. This numbering convention has become standard.

There are several mutations in the Pribnow box, two of which are shown in Figure 11-7, that prevent initiation of transcription. These mutations clearly indicate the importance of this sequence. Other bases outside of the Pribnow box are important too, as indicated by the other mutations shown in the figure.

Site Selection: III. The −35 Sequence

Examination of the complete sequence of the region protected by RNA polymerase indicates that for many (but not all) promoters, there is a second important region, to the left of the Pribnow box, whose sequences in different promoters have common features (Figure 11-8).

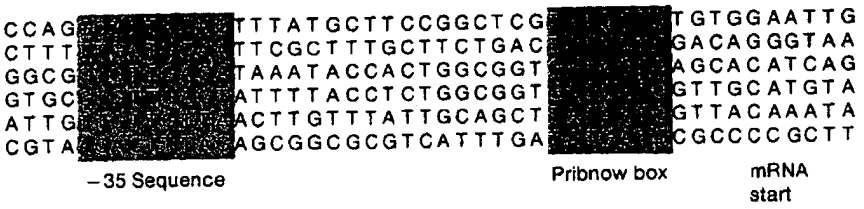


Figure 11-8
Base sequences in the noncoding strand of six different RNA polymerase-protected regions showing the similarity between the −35 sequences. In each case, mutations that eliminate promoter activity have been found in the −35 sequence. The vertical lines indicate the HindII cuts mentioned in the text. The Pribnow boxes rather than the mRNA start points are aligned.

This sequence, which is called the **−35 sequence** and typically contains nine bases, is thought to be the initial site of binding of the enzyme. Evidence for this notion comes from the following experiment. RNA polymerase is removed from the protected fragment and the fragment is purified. If fresh RNA polymerase is then added, binding will occur, indicating that the binding site is on the fragment. However, if the fragment is first treated with a restriction nuclease (Chapter 20) called HindII , which makes a double-strand break at the sites indicated in the figure by the lines, RNA polymerase can no longer bind; presumably, the binding site is destroyed by the nuclease. Thus, RNA polymerase is thought to bind first at the leftmost side of the protected region and then to the Pribnow box. How it moves from one site to the next is not known. A theory that the enzyme "slides" along the DNA was popular at one time; it has not been ruled out but is considered to be unlikely. Another possibility, which has some experimental support, is that the σ subunit binds first to a recognition site at the left in a highly specific interaction and then, owing to the great size of the enzyme, the appropriate region of the polymerase can come in contact with the Pribnow box region (Figure 11-9). Once bound to the Pribnow box, the polymerase then dissociates from the leftmost recognition site.

The **open-promoter complex** is a highly stable complex and is the active intermediate in chain initiation. In this complex a local unwinding ("melting") of the DNA helix occurs starting about ten base pairs from the left end of the Pribnow box and extending to the position of the first transcribed base. This melting is necessary for pairing of the incoming ribonucleotides. The base composition of the Pribnow box sequence (A+T-rich) renders the DNA strand susceptible to denaturation. Presumably RNA polymerase itself induces this conformational change.

The promoters discussed in this section are classified as **high-level** or **strong promoters.** There are also weak promoters in which recognition by RNA polymerase is poor. The number of RNA molecules

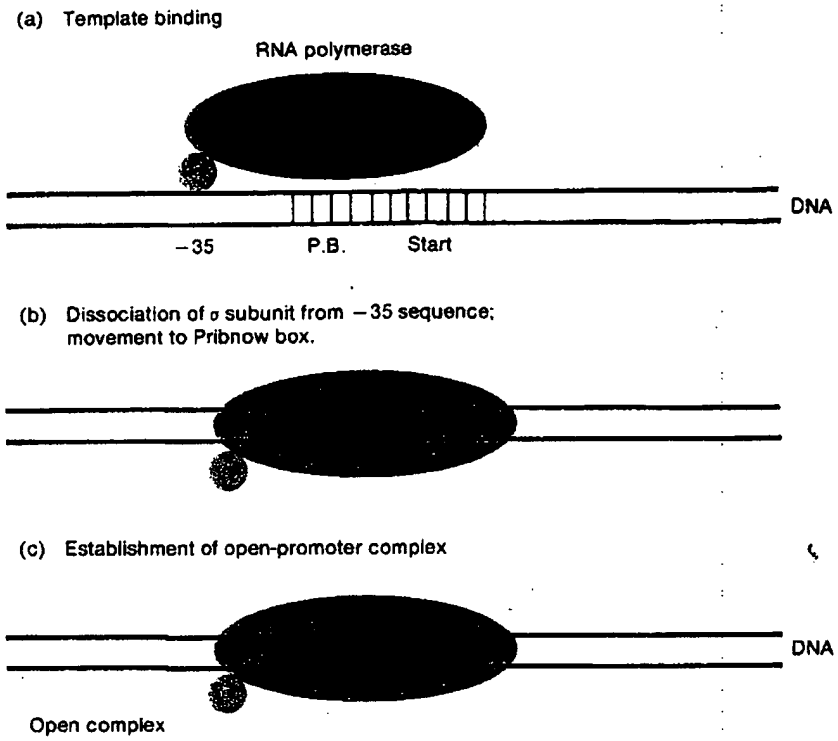


Figure 11-9
A proposed scheme for the binding of RNA polymerase to a promoter to form an open-promoter complex. Regions of the DNA molecule important for binding are shown in red. The shape of RNA polymerase is idealized for schematic purposes. The enzyme covers the region from bases −45 to +15, and the unpaired region in (c) extends from (roughly) −12 to +2. The enzyme is shown in contact with both strands because the strands are actually wrapped around one another in a helical array; however, true binding occurs only to bases in the coding strand. "P.B." indicates the Pribnow box.

synthesized per unit time from genes with weak promoters is much less than from a strong promoter with the result that fewer protein molecules are made per unit time by genes with weak promoters. Promoter strength is one factor which determines the number of copies of each protein molecule present in the cell. In most cases examined so far the difference between weak and strong promoters lies in the structure of the −35 region.

Site Selection: IV. The CAP Site

Some promoters totally lack the common −35 sequence—for example, the λ*pre*, *galP*, and *araBAD* promoters. These are active only in the presence of positive effector molecules (see Chapters 14 and 16); for example, the λ*pre* promoter is active only when the λ cII protein is present. The mechanisms of action of these effectors are not well understood, though a study of the *lac* promoter suggests that they bind to a site in the −50 to −30 region and, by a mechanism that differs from that